



(19)

Europäisches Patentamt

European Patent Office

Office européen des brevets



(11)

EP 1 266 015 B1

(12)

## EUROPEAN PATENT SPECIFICATION

(45) Date of publication and mention

of the grant of the patent:

30.08.2006 Bulletin 2006/35

(51) Int Cl.:

C12N 15/63 (2006.01)

C12N 15/65 (2006.01)

C12N 15/12 (2006.01)

C12N 15/62 (2006.01)

C07K 14/705 (2006.01)

C07K 14/72 (2006.01)

A01H 5/00 (2006.01)

C12Q 1/68 (2006.01)

G01N 33/68 (2006.01)

(21) Application number: 01922524.2

(22) Date of filing: 21.03.2001

(86) International application number:  
PCT/US2001/009050(87) International publication number:  
WO 2001/070816 (27.09.2001 Gazette 2001/39)

## (54) ECDYSONE RECEPTOR-BASED INDUCIBLE GENE EXPRESSION SYSTEM

INDUZIERTES GENEXPRESSIONSSYSTEM BASIEREND AUF ECDYSONREZEPTOREN

NOUVEAU SYSTEME D'EXPRESSION GENETIQUE INDUCIBLE FONDE SUR LE RECEPTEUR  
A L'ECDYSONE

(84) Designated Contracting States:

AT BE CH CY DE DK ES FI FR GB GR IE IT LI LU  
MC NL PT SE TR

- CHRISTOPHERSON K S ET AL: "ECDYSTEROID-DEPENDENT REGULATION OF GENES IN MAMMALIAN CELLS BY A DROSOPHILA ECDYSONE RECEPTOR AND CHIMERIC TRANSACTIVATORS" PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES OF USA, NATIONAL ACADEMY OF SCIENCE.

WASHINGTON, US, vol. 89, July 1992 (1992-07), pages 6314-6318, XP002921831 ISSN: 0027-8424

- KAKIZAWA TOMOKO ET AL: "Ligand-dependent heterodimerization of thyroid hormone receptor and retinoid X receptor." JOURNAL OF BIOLOGICAL CHEMISTRY, vol. 272, no. 38, 1997, pages 23799-23804, XP002200320 ISSN: 0021-9258

- PERERA SRINI C ET AL: "An analysis of ecdysone receptor domains required for heterodimerization with ultraspirel." ARCHIVES OF INSECT BIOCHEMISTRY AND PHYSIOLOGY, vol. 41, no. 2, 1999, pages 61-70, XP001077866 ISSN: 0739-4462

- KOELLE M R ET AL: "THE DROSOPHILA ECR GENE ENCODES AN ECDYSONE RECEPTOR, A NEW MEMBER OF THE STEROID RECEPTOR SUPERFAMILY" CELL, CELL PRESS, CAMBRIDGE, MA, US, vol. 67, no. 1, 4 October 1991 (1991-10-04), pages 59-77, XP001010154 ISSN: 0092-8674

(30) Priority: 22.03.2000 US 191355 P

20.02.2001 US 269799 P

(43) Date of publication of application:  
18.12.2002 Bulletin 2002/51(73) Proprietor: RheoGene Holdings Inc.  
Norristown, PA 19403 (US)

(72) Inventors:

- PALLI, Subba, Reddy  
Lansdale, PA 19446 (US)
- KAPITSKAYA, Marianna Zinovjevna  
North Wales, PA 19454 (US)
- CRESS, Dean, Ervin  
Souderton, PA 18964 (US)

(74) Representative: Harding, Charles Thomas et al  
D Young & Co  
120 Holborn  
London EC1N 2DY (GB)

(56) References cited:

EP-A- 0 798 378

EP-A- 0 965 644

WO-A-96/27673

WO-A-99/02683

WO-A-99/10510

WO-A-99/58155

Note: Within nine months from the publication of the mention of the grant of the European patent, any person may give notice to the European Patent Office of opposition to the European patent granted. Notice of opposition shall be filed in a written reasoned statement. It shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).

- LEID M ET AL: "PURIFICATION, CLONING, AND RXR IDENTITY OF THE HELA CELL FACTOR WITH WHICH RAR OR TR HETERODIMERIZES TO BIND TARGET SEQUENCES EFFICIENTLY" CELL, CELL PRESS, CAMBRIDGE, MA, US, vol. 68, 24 January 1992 (1992-01-24), pages 377-395, XP002007203 ISSN: 0092-8674
- PERERA SRINI C ET AL: "Studies on two ecdysone receptor isoforms of the spruce budworm, *Choristoneura fumiferana*." MOLECULAR AND CELLULAR ENDOCRINOLOGY, vol. 152, no. 1-2, 25 June 1999 (1999-06-25), pages 73-84, XP002200321 ISSN: 0303-7207

**Description****FIELD OF THE INVENTION**

[0001] This invention relates to the field of biotechnology or genetic engineering. Specifically, this invention relates to the field of gene expression. More specifically, this invention relates to a novel ecdysone receptor-based inducible gene expression system and methods of modulating the expression of a gene within a host cell using this inducible gene expression system.

**BACKGROUND OF THE INVENTION**

[0002] In the field of genetic engineering, precise control of gene expression is a valuable tool for studying, manipulating, and controlling development and other physiological processes. Gene expression is a complex biological process involving a number of specific protein-protein interactions. In order for gene expression to be triggered, such that it produces the RNA necessary as the first step in protein synthesis, a transcriptional activator must be brought into proximity of a promoter that controls gene transcription. Typically, the transcriptional activator itself is associated with a protein that has at least one DNA binding domain that binds to DNA binding sites present in the promoter regions of genes. Thus, for gene expression to occur, a protein comprising a DNA binding domain and a transactivation domain located at an appropriate distance from the DNA binding domain must be brought into the correct position in the promoter region of the gene.

[0003] The traditional transgenic approach utilizes a cell-type specific promoter to drive the expression of a designed transgene. A DNA construct containing the transgene is first incorporated into a host genome. When triggered by a transcriptional activator, expression of the transgene occurs in a given cell type.

[0004] Another means to regulate expression of foreign genes in cells is through inducible promoters. Examples of the use of such inducible promoters include the PR1-a promoter, prokaryotic repressor-operator systems, immunosuppressive-immunophilin systems, and higher eukaryotic transcription activation systems such as steroid hormone receptor systems and are described below.

[0005] The PR1-a promoter from tobacco is induced during the systemic acquired resistance response following pathogen attack. The use of PR1-a may be limited because it often responds to endogenous materials and external factors such as pathogens, UV-B radiation, and pollutants. Gene regulation systems based on promoters induced by heat shock, interferon and heavy metals have been described (Wurn et al., 1986, Proc. Natl. Acad. Sci. USA 83: 5414-5418; Arnheiter et al., 1990 Cell 62:51-61; Filmus et al., 1992 Nucleic Acids Research 20:27550-27560). However, these systems have limitations due to their effect on expression of non-target genes. These systems are also leaky.

[0006] Prokaryotic repressor-operator systems utilize bacterial repressor proteins and the unique operator DNA sequences to which they bind. Both the tetracycline ("Tet") and lactose ("Lac") repressor-operator systems from the bacterium *Escherichia coli* have been used in plants and animals to control gene expression. In the Tet system, tetracycline binds to the TetR repressor protein, resulting in a conformational change which releases the repressor protein from the operator which as a result allows transcription to occur. In the Lac system, a lac operon is activated in response to the presence of lactose, or synthetic analogs such as isopropyl-b-D-thiogalactoside. Unfortunately, the use of such systems is restricted by unstable chemistry of the ligands, i.e. tetracycline and lactose, their toxicity, their natural presence, or the relatively high levels required for induction or repression. For similar reasons, utility of such systems in animals is limited.

[0007] Immunosuppressive molecules such as FK506, rapamycin and cyclosporine A can bind to immunophilins FKBP12, cyclophilin, etc. Using this information, a general strategy has been devised to bring together any two proteins simply by placing FK506 on each of the two proteins or by placing FK506 on one and cyclosporine A on another one. A synthetic homodimer of FK506 (FK1012) or a compound resulted from fusion of FK506-cyclosporine (FKCsA) can then be used to induce dimerization of these molecules (Spencer et al., 1993, *Science* 262:1019-24; Belshaw et al., 1996 *Proc Natl Acad Sci U S A* 93:4604-7). Gal4 DNA binding domain fused to FKBP12 and VP16 activator domain fused to cyclophilin, and FKCAs compound were used to show heterodimerization and activation of a reporter gene under the control of a promoter containing Gal4 binding sites. Unfortunately, this system includes immunosuppressants that can have unwanted side effects and therefore, limits its use for various mammalian gene switch applications.

[0008] Higher eukaryotic transcription activation systems such as steroid hormone receptor systems have also been employed. Steroid hormone receptors are members of the nuclear receptor superfamily and are found in vertebrate and invertebrate cells. Unfortunately, use of steroidal compounds that activate the receptors for the regulation of gene expression, particularly in plants and mammals, is limited due to their involvement in many other natural biological pathways in such organisms. In order to overcome such difficulties, an alternative system has been developed using insect ecdysone receptors (EcR).

[0009] Growth, molting, and development in insects are regulated by the ecdysone steroid hormone (molting hormone)

and the juvenile hormones (Dhadialla, et al., 1998. Annu. Rev. Entomol. 43: 545-569). The molecular target for ecdysone in insects consists of at least ecdysone receptor (EcR) and ultraspiracle protein (USP). EcR is a member of the nuclear steroid receptor super family that is characterized by signature DNA and ligand binding domains, and an activation domain (Koelle et al. 1991, Cell, 67:59-77). EcR receptors are responsive to a number of steroid compounds such as 5 ponasterone A and muristerone A. Recently, non-steroidal compounds with ecdysteroid agonist activity have been described, including the commercially available insecticides tebufenozide and methoxyfenozide that are marketed world wide by Rohm and Haas Company (see International Patent Application No. PCT/EP96/00686 (WO 96/27673) and US Patent 5,530,028). Both analogs have exceptional safety profiles to other organisms.

[0010] International Patent Application No. PCT/US97/05330 (WO 97/38117) discloses methods for modulating the expression of an exogenous gene in which a DNA construct comprising the exogenous gene and an ecdysone response element is activated by a second DNA construct comprising an ecdysone receptor that, in the presence of a ligand therefor, and optionally in the presence of a receptor capable of acting as a silent partner, binds to the ecdysone response element to induce gene expression. The ecdysone receptor of choice was isolated from *Drosophila melanogaster*. Typically, such systems require the presence of the silent partner, preferably retinoid X receptor (RXR), in order to provide optimum activation. In mammalian cells, insect ecdysone receptor (EcR) heterodimerizes with retinoid X receptor (RXR) and regulates expression of target genes in a ligand dependent manner. International Patent Application No. PCT/US98/14215 (WO 99/02683) discloses that the ecdysone receptor isolated from the silk moth *Boimbyx mori* is functional in mammalian systems without the need for an exogenous dimer partner.

[0011] U.S. Patent No. 5,880,333 discloses a *Drosophila melanogaster* EcR and ultraspiracle (USP) heterodimer system used in plants in which the transactivation domain and the DNA binding domain are positioned on two different hybrid proteins. Unfortunately, this system is not effective for inducing reporter gene expression in animal cells (for comparison, see Example 1.2, below).

[0012] In each of these cases, the transactivation domain and the DNA binding domain (either as native EcR as in International Patent Application No. PCT/US98/14215 or as modified EcR as in International Patent Application No. PCT/US97/05330) were incorporated into a single molecule and the other heterodimeric partners, either USP or RXR, were used in their full length form.

[0013] Drawbacks of the above described EcR-based gene regulation systems include a considerable background activity in the absence of ligands and that these systems are not applicable for use in both plants and animals (see U.S. Patent No. 5,880,333). For most applications that rely on modulating gene expression, these EcR-based systems are undesirable. Therefore, a need exists in the art for improved systems to precisely modulate the expression of exogenous genes in both plants and animals. Such improved systems would be useful for applications such as gene therapy, large scale production of proteins and antibodies, cell-based high throughput screening assays, functional genomics and regulation of traits in transgenic animals. Improved systems that are simple, compact, and dependent on ligands that are relatively inexpensive, readily available, and of low toxicity to the host would prove useful for regulating biological systems.

[0014] Various publications are cited herein, the disclosures of which are incorporated by reference in their entireties. However, the citation of any reference herein should not be construed as an admission that such reference is available as "Prior Art" to the instant application.

#### 40 SUMMARY OF THE INVENTION

[0015] The present invention relates to a novel ecdysone receptor-based inducible gene expression system, novel receptor polynucleotides and polypeptides for use in the novel inducible gene expression system, and methods of modulating the expression of a gene within a host cell using this inducible gene expression system. In particular, Applicants' invention relates to an improved gene expression modulation system comprising a polynucleotide encoding a receptor polypeptide comprising a truncation mutation.

[0016] Specifically, the present invention relates to a gene expression modulation system comprising: a) a first gene expression cassette that is capable of being expressed in a host cell comprising a polynucleotide that encodes a first polypeptide comprising: i) a DNA-binding domain that recognizes a response element associated with a gene whose expression is to be modulated; and ii) a ligand binding domain comprising a ligand binding domain from a nuclear receptor; and b) a second gene expression cassette that is capable of being expressed in the host cell comprising a polynucleotide sequence that encodes a second polypeptide comprising: i) a transactivation domain; and ii) a ligand binding domain comprising a ligand binding domain from a nuclear receptor other than an ultraspiracle receptor; wherein the DNA binding domain and the transactivation domain are from a polypeptide other than an ecdysone receptor, a retinoid X receptor, or an ultraspiracle receptor, wherein the ligand binding domains from the first polypeptide and the second polypeptide are different and dimerize.

[0017] In a specific embodiment, the ligand binding domain of the first polypeptide comprises an ecdysone receptor (EcR) ligand binding domain

[0018] In another specific embodiment, the ligand binding domain of the second polypeptide comprises a retinoid X receptor (RXR) ligand binding domain.

[0019] In a preferred embodiment, the ligand binding domain of the first polypeptide comprises an ecdysone receptor ligand binding domain and the ligand binding domain of the second polypeptide comprises a retinoid X receptor ligand binding domain

[0020] The present invention also relates to a gene expression modulation system according to the invention further comprising c) a third gene expression cassette comprising: i) a response element to which the DNA-binding domain of the first polypeptide binds; ii) a promoter that is activated by the transactivation domain of the second polypeptide; and iii) the gene whose expression is to be modulated.

[0021] The present invention also relates to an isolated polynucleotide encoding a truncated EcR or a truncated RXR polypeptide, wherein the truncation mutation affects ligand binding activity or ligand sensitivity.

[0022] In particular, the present invention relates to an isolated polynucleotide encoding a truncated EcR or a truncated RXR polypeptide comprising a truncation mutation that reduces ligand binding activity or ligand sensitivity of said EcR or RXR polypeptide. In a specific embodiment, the present invention relates to an isolated polynucleotide encoding a truncated EcR or a truncated RXR polypeptide comprising a truncation mutation that reduces steroid binding activity or steroid sensitivity of said EcR or RXR polypeptide. In another specific embodiment, the present invention relates to an isolated polynucleotide encoding a truncated EcR or a truncated RXR polypeptide comprising a truncation mutation that reduces non-steroid binding activity or non-steroid sensitivity of said EcR or RXR polypeptide.

[0023] The present invention also relates to an isolated polynucleotide encoding a truncated EcR or a truncated RXR polypeptide comprising a truncation mutation that enhances ligand binding activity or ligand sensitivity of said EcR or RXR polypeptide. In a specific embodiment, the present invention relates to an isolated polynucleotide encoding a truncated EcR or a truncated RXR polypeptide comprising a truncation mutation that enhances steroid binding activity or steroid sensitivity of said EcR or RXR polypeptide. In another specific embodiment, the present invention relates to an isolated polynucleotide encoding a truncated EcR or a truncated RXR polypeptide comprising a truncation mutation that enhances non-steroid binding activity or non-steroid sensitivity of said EcR or RXR polypeptide.

[0024] The present invention also relates to an isolated polynucleotide encoding a truncated RXR polypeptide comprising a truncation mutation that increases ligand sensitivity of a heterodimer comprising the truncated retinoid X receptor polypeptide and a dimerization partner. In a specific embodiment, the dimerization partner is an ecdysone receptor polypeptide.

[0025] The present invention also relates to an isolated polypeptide encoded by a polynucleotide according to Applicants' invention. In particular, the present invention relates to an isolated truncated EcR or truncated RXR polypeptide comprising a truncation mutation, wherein the EcR or RXR polypeptide is encoded by a polynucleotide according to the invention.

[0026] Thus, the present invention also relates to an isolated truncated EcR or truncated RXR polypeptide comprising a truncation mutation that affects ligand binding activity or ligand sensitivity of said EcR or RXR polypeptide.

[0027] Applicants' invention also relates to methods of modulating gene expression in a host cell using a gene expression modulation system according to the invention. Specifically, Applicants' invention provides a method of modulating the expression of a gene in a host cell comprising the gene to be modulated comprising the steps of: a) introducing into the host cell a gene expression modulation system according to the invention; and b) introducing into the host cell a ligand that independently combines with the ligand binding domains of the first polypeptide and the second polypeptide of the gene expression modulation system; wherein the gene to be expressed is a component of a chimeric gene comprising: i) a response element comprising a domain to which the DNA binding domain from the first polypeptide binds; ii) a promoter that is activated by the transactivation domain of the second polypeptide; and iii) the gene whose expression is to be modulated, whereby a complex is formed comprising the ligand, the first polypeptide, and the second polypeptide, and whereby the complex modulates expression of the gene in the host cell.

[0028] Applicants' invention also provides an isolated host cell comprising an inducible gene expression system according to the invention. The present invention also relates to an isolated host cell comprising a polynucleotide or polypeptide according to the invention. Accordingly, Applicants' invention also relates to a non-human organism comprising a host cell according to the invention.

#### BRIEF DESCRIPTION OF THE DRAWINGS

[0029]

Figure 1: An ecdysone receptor-based gene expression system comprising a first gene expression cassette encoding a Gal4DBD-CfEcRDEF chimeric polypeptide and a second gene expression cassette encoding a VP16AD-MmRXR-DEF chimeric polypeptide; prepared as described in Example 1 (switch 1.1).

Figure 2: An ecdysone receptor-based gene expression system comprising a first gene expression cassette encoding

a Gal4DBD-CfEcRDEF chimeric polypeptide and a second gene expression cassette encoding a VP16AD-CIUSP-DEF chimeric polypeptide; prepared as described in Example 1 (switch 1.2).

Figure 3: An ecdysone receptor-based gene expression system comprising a first gene expression cassette encoding a Gal4DBD-MmRXRDEF chimeric polypeptide and a second gene expression cassette encoding a VP16AD-CfEcRCDEF chimeric polypeptide; prepared as described in Example 1 (switch 1.3).

Figure 4: An ecdysone receptor-based gene expression system comprising a first gene expression cassette encoding a Gal4DBD-MmRXRDEF chimeric polypeptide and a second gene expression cassette encoding a VP16AD-DmEcRCDEF chimeric polypeptide; prepared as described in Example 1 (switch 1.4).

Figure 5: An ecdysone receptor-based gene expression system comprising a first gene expression cassette encoding a Gal4DBD-CfUASPDEF chimeric polypeptide and a second gene expression cassette encoding a VP16AD-CfEcRCDEF chimeric polypeptide; prepared as described in Example 1 (switch 1.5).

Figure 6: An ecdysone receptor-based gene expression system comprising a first gene expression cassette encoding a Gal4DBD-CfEcRDEF-VP16AD chimeric polypeptide; prepared as described in Example 1 (switch 1.6).

Figure 7: An ecdysone receptor-based gene expression system comprising a first gene expression cassette encoding a VP16AD-CfEcRCDEF chimeric polypeptide; prepared as described in Example 1 (switch 1.7).

Figure 8: An ecdysone receptor-based gene expression system comprising a first gene expression cassette encoding a VP16AD-DmEcRCDEF chimeric polypeptide and a second gene expression cassette encoding a MmRXR polypeptide; prepared as described in Example 1 (switch 1.8).

Figure 9: An ecdysone receptor-based gene expression system comprising a first gene expression cassette encoding a VP16AD-CfEcRCDEF chimeric polypeptide and a second gene expression cassette encoding a MmRXR polypeptide; prepared as described in Example 1 (switch 1.9).

Figure 10: An ecdysone receptor-based gene expression system comprising a gene expression cassette encoding a Gal4DBD-CfEcRCDEF chimeric polypeptide; prepared as described in Example 1 (switch 1.10).

Figure 11: Expression data of GAL4CfEcRA/BCDEF, GAL4CfEcRCDEF, GAL4CfEcR1/2CDEF, GAL4CfEcRDEF, GAL4CfEcREF, GAL4CfEcRDE truncation mutants transfected into NIH3T3 cells along with VP16MmRXRDE, pFRLUC and pTKRL plasmid DNAs.

Figure 12: Expression data of GAL4CfEcRA/BCDEF, GAL4CfEcRCDEF, GAL4CfEcR1/2CDEF, GAL4CfEcRDEF, GAL4CfEcREF, GAL4CfEcRDE truncation mutants transfected into 3T3 cells along with VP16MmRXRE, pFRLUC and pTKRL plasmid DNAs.

Figure 13: Expression data of VP16MmRXRA/BCDEF, VP16MmRXRCDEF, VP16MmRXRDEF, VP16MmRXREF, VP16MmRXRBam-EF, VP16MmRXRAF2del constructs transfected into NIH3T3 cells along with GALACfEcRCDEF, pFRLUC and pTKRL plasmid DNAs.

Figure 14: Expression data of VP16MmRXRA/BCDEF, VP16MmRXRCDEF, VP16MmRXRDEF, VP16MmRXREF, VP16MmRXRBam-EF, VP16MmRXRAF2del constructs transfected into NLH3T3 cells along with GALACfEcRDEF, pFRLUC and pTKRL plasmid DNAs.

Figure 15: Expression data of various truncated CfEcR and MmRXR receptor pairs transfected into NIH3T3 cells along with GALACfEcRDEF, pFRLUC and pTKRL plasmid DNAs.

#### DETAILED DESCRIPTION OF THE INVENTION

**[0030]** Applicants have now developed an improved ecdysone receptor-based inducible gene expression system comprising a truncation mutant of an ecdysone receptor or a retinoid X receptor (RXR) polypeptide that affects ligand binding activity or ligand sensitivity. This mutational effect may increase or reduce ligand binding activity or ligand sensitivity and may be steroid or non-steroid specific. Thus, Applicants' invention provides an improved ecdysone receptor-based inducible gene expression system useful for modulating expression of a gene of interest in a host cell. In a particularly desirable embodiment, Applicants' invention provides an inducible gene expression system that has a reduced level of background gene expression and responds to submicromolar concentrations of non-steroidal ligand. Thus, Applicants' novel inducible gene expression system and its use in methods of modulating gene expression in a host cell overcome the limitations of currently available inducible expression systems and provide the skilled artisan with an effective means to control gene expression.

**[0031]** The present invention provides a novel inducible gene expression system that can be used to modulate gene expression in both prokaryotic and eukaryotic host cells. Applicants' invention is useful for applications such as gene therapy, large scale production of proteins and antibodies, cell-based high throughput screening assays, functional genomics and regulation of traits in transgenic organisms.

#### DEFINITIONS

**[0032]** In this disclosure, a number of terms and abbreviations are used. The following definitions are provided and

should be helpful in understanding the scope and practice of the present invention.

[0033] In a specific embodiment, the term "about" or "approximately" means within 20%, preferably within 10%, more preferably within 5%, and even more preferably within 1% of a given value or range.

[0034] The term "substantially free" means that a composition comprising "A" (where "A" is a single protein, DNA molecule, vector, recombinant host cell, etc.) is substantially free of "B" (where "B" comprises one or more contaminating proteins, DNA molecules, vectors, etc.) when at least about 75% by weight of the proteins, DNA, vectors (depending on the category of species to which A and B belong) in the composition is "A". Preferably, "A" comprises at least about 90% by weight of the A+B species in the composition, most preferably at least about 99% by weight. It is also preferred that a composition, which is substantially free of contamination, contain only a single molecular weight species having the activity or characteristic of the species of interest.

[0035] The term "isolated" for the purposes of the present invention designates a biological material (nucleic acid or protein) that has been removed from its original environment (the environment in which it is naturally present).

[0036] For example, a polynucleotide present in the natural state in a plant or an animal is not isolated. The same polynucleotide separated from the adjacent nucleic acids in which it is naturally present. The term "purified" does not require the material to be present in a form exhibiting absolute purity, exclusive of the presence of other compounds. It is rather a relative definition.

[0037] A polynucleotide is in the "purified" state after purification of the starting material or of the natural material by at least one order of magnitude, preferably 2 or 3 and preferably 4 or 5 orders of magnitude.

[0038] A "nucleic acid" is a polymeric compound comprised of covalently linked subunits called nucleotides. Nucleic acid includes polyribonucleic acid (RNA) and polydeoxyribonucleic acid (DNA), both of which may be single-stranded or double-stranded. DNA includes but is not limited to cDNA, genomic DNA, plasmids DNA, synthetic DNA, and semi-synthetic DNA. DNA may be linear, circular, or supercoiled.

[0039] A "nucleic acid molecule" refers to the phosphate ester polymeric form of nbonucleosides (adenosine, guanosine, uridine or cytidine; "RNA molecules") or deoxyribonucleosides (deoxyadenosine, deoxyguanosine, deoxythymidine, or deoxycytidine; "DNA molecules"), or any phosphoester analogs thereof, such as phosphorothioates and thioesters, in either single stranded form, or a double-stranded helix. Double stranded DNA-DNA, DNA-RNA and RNA-RNA helices are possible. The term nucleic acid molecule, and in particular DNA or RNA molecule, refers only to the primary and secondary structure of the molecule, and does not limit it to any particular tertiary forms. Thus, this term includes double-stranded DNA found, *inter alia*, in linear or circular DNA molecules (*e.g.*, restriction fragments), plasmids, and chromosomes. In discussing the structure of particular double-stranded DNA molecules, sequences may be described herein according to the normal convention of giving only the sequence in the 5' to 3' direction along the non-transcribed strand of DNA (*i.e.*, the strand having a sequence homologous to the mRNA). A "recombinant DNA molecule" is a DNA molecule that has undergone a molecular biological manipulation.

[0040] The term "fragment" will be understood to mean a nucleotide sequence of reduced length relative to the reference nucleic acid and comprising, over the common portion, a nucleotide sequence identical to the reference nucleic acid. Such a nucleic acid fragment according to the invention may be, where appropriate, included in a larger polynucleotide of which it is a constituent. Such fragments comprise, or alternatively consist of, oligonucleotides ranging in length from at least 8, 10, 12, 15, 18, 20 to 25, 30, 40, 50, 70, 80, 100, 200, 500, 1000 or 1500 consecutive nucleotides of a nucleic acid according to the invention.

[0041] As used herein, an "isolated nucleic acid fragment" is a polymer of RNA or DNA that is single- or double-stranded, optionally containing synthetic, non-natural or altered nucleotide bases. An isolated nucleic acid fragment in the form of a polymer of DNA may be comprised of one or more segments of cDNA, genomic DNA or synthetic DNA.

[0042] A "gene" refers to an assembly of nucleotides that encode a polypeptide, and includes cDNA and genomic DNA nucleic acids. "Gene" also refers to a nucleic acid fragment that expresses a specific protein or polypeptide, including regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence. "Native gene" refers to a gene as found in nature with its own regulatory sequences. "Chimeric gene" refers to any gene that is not a native gene, comprising regulatory and/or coding sequences that are not found together in nature. Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature. A chimeric gene may comprise coding sequences derived from different sources and/or regulatory sequences derived from different sources. "Endogenous gene" refers to a native gene in its natural location in the genome of an organism. A "foreign" gene or "heterologous" gene refers to a gene not normally found in the host organism, but that is introduced into the host organism by gene transfer. Foreign genes can comprise native genes inserted into a non-native organism, or chimeric genes. A "transgene" is a gene that has been introduced into the genome by a transformation procedure.

[0043] "Heterologous" DNA refers to DNA not naturally located in the cell, or in a chromosomal site of the cell. Preferably, the heterologous DNA includes a gene foreign to the cell.

[0044] The term "genome" includes chromosomal as well as mitochondrial, chloroplast and viral DNA or RNA.

**[0045]** A nucleic acid molecule is "hybridizable" to another nucleic acid molecule, such as a cDNA, genomic DNA, or RNA, when a single stranded form of the nucleic acid molecule can anneal to the other nucleic acid molecule under the appropriate conditions of temperature and solution ionic strength (see *Sambrook et al.*, 1989 *infra*). Hybridization and washing conditions are well known and exemplified in Sambrook, J., Fritsch, E. F. and Maniatis, T. *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (1989), particularly Chapter 11 and Table 11.1 therein (entirely incorporated herein by reference). The conditions of temperature and ionic strength determine the "stringency" of the hybridization.

**[0046]** Stringency conditions can be adjusted to screen for moderately similar fragments, such as homologous sequences from distantly related organisms, to highly similar fragments, such as genes that duplicate functional enzymes from closely related organisms. For preliminary screening for homologous nucleic acids, low stringency hybridization conditions, corresponding to a  $T_m$  of 55°, can be used, e.g., 5x SSC, 0.1% SDS, 0.25% milk, and no formamide; or 30% formamide, 5x SSC, 0.5% SDS). Moderate stringency hybridization conditions correspond to a higher  $T_m$ , e.g., 40% formamide, with 5x or 6x SCC. High stringency hybridization conditions correspond to the highest  $T_m$ , e.g., 50% formamide, 5x or 6x SCC. Hybridization requires that the two nucleic acids contain complementary sequences, although depending on the stringency of the hybridization, mismatches between bases are possible.

**[0047]** The term "complementary" is used to describe the relationship between nucleotide bases that are capable of hybridizing to one another. For example, with respect to DNA, adenosine is complementary to thymine and cytosine is complementary to guanine. Accordingly, the instant invention also includes isolated nucleic acid fragments that are complementary to the complete sequences as disclosed or used herein as well as those substantially similar nucleic acid sequences.

**[0048]** In a specific embodiment, the term "standard hybridization conditions" refers to a  $T_m$  of 55°C, and utilizes conditions as set forth above. In a preferred embodiment, the  $T_m$  is 60°C; in a more preferred embodiment, the  $T_m$  is 65 °C.

**[0049]** Post-hybridization washes also determine stringency conditions. One set of preferred conditions uses a series of washes starting with 6X SSC, 0.5% SDS at room temperature for 15 minutes (min), then repeated with 2X SSC, 0.5% SDS at 45°C for 30 minutes, and then repeated twice with 0.2X SSC, 0.5% SDS at 50°C for 30 minutes. A more preferred set of stringent conditions uses higher temperatures in which the washes are identical to those above except for the temperature of the final two 30 min washes in 0.2X SSC, 0.5% SDS was increased to 60°C. Another preferred set of highly stringent conditions uses two final washes in 0.1X SSC, 0.1% SDS at 65°C. Hybridization requires that the two nucleic acids comprise complementary sequences, although depending on the stringency of the hybridization, mismatches between bases are possible.

**[0050]** The appropriate stringency for hybridizing nucleic acids depends on the length of the nucleic acids and the degree of complementation, variables well known in the art. The greater the degree of similarity or homology between two nucleotide sequences, the greater the value of  $T_m$  for hybrids of nucleic acids having those sequences. The relative stability (corresponding to higher  $T_m$ ) of nucleic acid hybridizations decreases in the following order: RNA:RNA; DNA:RNA, DNA:DNA. For hybrids of greater than 100 nucleotides in length, equations for calculating  $T_m$  have been derived (see Sambrook *et al.*, *supra*, 9.50-0.51). For hybridization with shorter nucleic acids, i.e., oligonucleotides, the position of mismatches becomes more important, and the length of the oligonucleotide determines its specificity (see Sambrook *et al.*, *supra*, 11.7-11.8).

**[0051]** In one embodiment the length for a hybridizable nucleic acid is at least about 10 nucleotides. Preferable a minimum length for a hybridizable nucleic acid is at least about 15 nucleotides; more preferably at least about 20 nucleotides; and most preferably the length is at least 30 nucleotides. Furthermore, the skilled artisan will recognize that the temperature and wash solution salt concentration may be adjusted as necessary according to factors such as length of the probe.

**[0052]** The term "probe" refers to a single-stranded nucleic acid molecule that can base pair with a complementary single stranded target nucleic acid to form a double-stranded molecule.

**[0053]** As used herein, the term "oligonucleotide" refers to a nucleic acid, generally of at least 18 nucleotides, that is hybridizable to a genomic DNA molecule, a cDNA molecule, a plasmid DNA or an mRNA molecule. Oligonucleotides can be labeled, e. g., with  $^{32}P$ -nucleotides or nucleotides to which a label, such as biotin, has been covalently conjugated. A labeled oligonucleotide can be used as a probe to detect the presence of a nucleic acid. Oligonucleotides (one or both of which may be labeled) can be used as PCR primers, either for cloning full length or a fragment of a nucleic acid, or to detect the presence of a nucleic acid. An oligonucleotide can also be used to form a triple helix with a DNA molecule. Generally, oligonucleotides are prepared synthetically, preferably on a nucleic acid synthesizer. Accordingly, oligonucleotides can be prepared with non-naturally occurring phosphoester analog bonds, such as thioester bonds, etc.

**[0054]** A "primer" is an oligonucleotide that hybridizes to a target nucleic acid sequence to create a double stranded nucleic acid region that can serve as an initiation point for DNA synthesis under suitable conditions. Such primers may be used in a polymerase chain reaction.

**[0055]** "Polymerase chain reaction" is abbreviated PCR and means an *in vitro* method for enzymatically amplifying specific nucleic acid sequences. PCR involves a repetitive series of temperature cycles with each cycle comprising three

stages: denaturation of the template nucleic acid to separate the strands of the target molecule, annealing a single stranded PCR oligonucleotide primer to the template nucleic acid, and extension of the annealed primer(s) by DNA polymerase. PCR provides a means to detect the presence of the target molecule and, under quantitative or semi-quantitative conditions, to determine the relative amount of that target molecule within the starting pool of nucleic acids.

**[0056]** "Reverse transcription-polymerase chain reaction" is abbreviated RT-PCR and means an *in vitro* method for enzymatically producing a target cDNA molecule or molecules from an RNA molecule or molecules, followed by enzymatic amplification of a specific nucleic acid sequence or sequences within the target cDNA molecule or molecules as described above. RT-PCR also provides a means to detect the presence of the target molecule and, under quantitative or semi-quantitative conditions, to determine the relative amount of that target molecule within the starting pool of nucleic acids.

**[0057]** A DNA "coding sequence" is a double-stranded DNA sequence that is transcribed and translated into a polypeptide in a cell *in vitro* or *in vivo* when placed under the control of appropriate regulatory sequences. "Suitable regulatory sequences" refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences may include promoters, translation leader sequences, introns,

polyadenylation recognition sequences, RNA processing site, effector binding site and stem-loop structure. The boundaries of the coding sequence are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxyl) terminus. A coding sequence can include, but is not limited to, prokaryotic sequences, cDNA from mRNA, genomic DNA sequences, and even synthetic DNA sequences. If the coding sequence is intended for expression in a eukaryotic cell, a polyadenylation signal and transcription termination sequence will usually be located 3' to the coding sequence.

**[0058]** "Open reading frame" is abbreviated ORF and means a length of nucleic acid sequence, either DNA, cDNA or RNA, that comprises a translation start signal or initiation codon, such as an ATG or AUG, and a termination codon and can be potentially translated into a polypeptide sequence.

**[0059]** The term "head-to-head" is used herein to describe the orientation of two polynucleotide sequences in relation to each other. Two polynucleotides are positioned in a head-to-head orientation when the 5' end of the coding strand of one polynucleotide is adjacent to the 5' end of the coding strand of the other polynucleotide, whereby the direction of transcription of each polynucleotide proceeds away from the 5' end of the other polynucleotide. The term "head-to-head" may be abbreviated (5')-to-(5') and may also be indicated by the symbols ( $\leftarrow$ ) or (3' $\leftarrow$ 5' $\rightarrow$ 3').

**[0060]** The term "tail-to-tail" is used herein to describe the orientation of two polynucleotide sequences in relation to each other. Two polynucleotides are positioned in a tail-to-tail orientation when the 3' end of the coding strand of one polynucleotide is adjacent to the 3' end of the coding strand of the other polynucleotide, whereby the direction of transcription of each polynucleotide proceeds toward the other polynucleotide. The term "tail-to-tail" may be abbreviated (3')-to-(3') and may also be indicated by the symbols ( $\rightarrow$ ) or (5' $\rightarrow$ 3' $\leftarrow$ 5').

**[0061]** The term "head-to-tail" is used herein to describe the orientation of two polynucleotide sequences in relation to each other. Two polynucleotides are positioned in a head-to-tail orientation when the 5' end of the coding strand of one polynucleotide is adjacent to the 3' end of the coding strand of the other polynucleotide, whereby the direction of transcription of each polynucleotide proceeds in the same direction as that of the other polynucleotide. The term "head-to-tail" may be abbreviated (5')-to-(3') and may also be indicated by the symbols ( $\rightarrow$ ) or (5' $\rightarrow$ 3' $\rightarrow$ 3').

**[0062]** The term "downstream" refers to a nucleotide sequence that is located 3' to reference nucleotide sequence. In particular, downstream nucleotide sequences generally relate to sequences that follow the starting point of transcription. For example, the translation initiation codon of a gene is located downstream of the start site of transcription.

**[0063]** The term "upstream" refers to a nucleotide sequence that is located 5' to reference nucleotide sequence. In particular, upstream nucleotide sequences generally relate to sequences that are located on the 5' side of a coding sequence or starting point of transcription. For example, most promoters are located upstream of the start site of transcription.

**[0064]** The terms "restriction endonuclease" and "restriction enzyme" refer to an enzyme that binds and cuts within a specific nucleotide sequence within double stranded DNA.

**[0065]** "Homologous recombination" refers to the insertion of a foreign DNA sequence into another DNA molecule, e.g., insertion of a vector in a chromosome. Preferably, the vector targets a specific chromosomal site for homologous recombination. For specific homologous recombination, the vector will contain sufficiently long regions of homology to sequences of the chromosome to allow complementary binding and incorporation of the vector into the chromosome. Longer regions of homology, and greater degrees of sequence similarity, may increase the efficiency of homologous recombination.

**[0066]** Several methods known in the art may be used to propagate a polynucleotide according to the invention. Once a suitable host system and growth conditions are established, recombinant expression vectors can be propagated and prepared in quantity. As described herein, the expression vectors which can be used include, but are not limited to, the following vectors or their derivatives: human or animal viruses such as vaccinia virus or adenovirus; insect viruses such as baculovirus; yeast vectors; bacteriophage vectors (e.g., lambda), and plasmid and cosmid DNA vectors, to name but

a few.

[0067] A "vector" is any means for the cloning of and/or transfer of a nucleic acid into a host cell. A vector may be a replicon to which another DNA segment may be attached so as to bring about the replication of the attached segment. A "replicon" is any genetic element (e.g., plasmid, phage, cosmid, chromosome, virus) that functions as an autonomous unit of DNA replication *in vivo*, i.e., capable of replication under its own control. The term "vector" includes both viral and nonviral means for introducing the nucleic acid into a cell *in vitro*, *ex vivo* or *in vivo*. A large number of vectors known in the art may be used to manipulate nucleic acids, incorporate response elements and promoters into genes, etc. Possible vectors include, for example, plasmids or modified viruses including, for example bacteriophages such as lambda derivatives, or plasmids such as PBR322 or pUC plasmid derivatives, or the Bluescript vector. For example, the insertion of the DNA fragments corresponding to response elements and promoters into a suitable vector can be accomplished by ligating the appropriate DNA - fragments into a chosen vector that has complementary cohesive termini. Alternatively, the ends of the DNA molecules may be enzymatically modified or any site may be produced by ligating nucleotide sequences (linkers) into the DNA termini. Such vectors may be engineered to contain selectable marker genes that provide for the selection of cells that have incorporated the marker into the cellular genome. Such markers allow identification and/or selection of host cells that incorporate and express the proteins encoded by the marker.

[0068] Viral vectors, and particularly retroviral vectors, have been used in a wide variety of gene delivery applications in cells, as well as living animal subjects. Viral vectors that can be used include but are not limited to retroviruses, adeno-associated virus, pox, baculovirus, vaccinia, herpes simplex, Epstein-Barr, adenovirus, geminivirus, and cauliimyovirus vectors. Non-viral vectors include plasmids, liposomes, electrically charged lipids (cytotoxins), DNA-protein complexes, and biopolymers. In addition to a nucleic acid, a vector may also comprise one or more regulatory regions, and/or selectable markers useful in selecting, measuring, and monitoring nucleic acid transfer results (transfer to which tissues, duration of expression, etc.).

[0069] The term "plasmid" refers to an extra chromosomal element often carrying a gene that is not part of the central metabolism of the cell, and usually in the form of circular double-stranded DNA molecules. Such elements may be autonomously replicating sequences, genome integrating sequences, phage or nucleotide sequences, linear, circular, or supercoiled, of a single- or double-stranded DNA or RNA, derived from any source, in which a number of nucleotide sequences have been joined or recombined into a unique construction which is capable of introducing a promoter fragment and DNA sequence for a selected gene product along with appropriate 3' untranslated sequence into a cell.

[0070] A "cloning vector" is a "replicon", which is a unit length of a nucleic acid, preferably DNA, that replicates sequentially and which comprises an origin of replication, such as a plasmid, phage or cosmid, to which another nucleic acid segment may be attached so as to bring about the replication of the attached segment. Cloning vectors may be capable of replication in one cell type and expression in another ("shuttle vector").

[0071] Vectors may be introduced into the desired host cells by methods known in the art, e.g., transfection, electroporation, microinjection, transduction, cell fusion, DEAE dextran, calcium phosphate precipitation, lipofection (lysosome fusion), use of a gene gun, or a DNA vector transporter (see, e.g., Wu et al., 1992, J. Biol. Chem. 267:963-967; Wu and Wu, 1988, J. Biol. Chem. 263:14621-14624; and Hartmut et al., Canadian Patent Application No. 2,012,311, filed March 15, 1990).

[0072] A polynucleotide according to the invention can also be introduced *in vivo* by lipofection. For the past decade, there has been increasing use of liposomes for encapsulation and transfection of nucleic acids *in vitro*. Synthetic cationic lipids designed to limit the difficulties and dangers encountered with liposome mediated transfection can be used to prepare liposomes for *in vivo* transfection of a gene encoding a marker (Felper et al., 1987, PNAS 84:7413; Mackey, et al., 1988, Proc. Natl. Acad. Sci. U.S.A. 85:8027-8031; and Ulmer et al., 1993, Science 259:1745-1748). The use of cationic lipids may promote encapsulation of negatively charged nucleic acids, and also promote fusion with negatively charged cell membranes (Felgner and Ringold, 1989, Science 337:387-388). Particularly useful lipid compounds and compositions for transfer of nucleic acids are described in International Patent Publications WO95/18863 and WO95/17823, and in U.S. Patent No. 5,459,127. The use of lipofection to introduce exogenous genes into the specific organs *in vivo* has certain practical advantages. Molecular targeting of liposomes to specific cells represents one area of benefit. It is clear that directing transfection to particular cell types would be particularly preferred in a tissue with cellular heterogeneity, such as pancreas, liver, kidney, and the brain. Lipids may be chemically coupled to other molecules for the purpose of targeting (Mackey, et al., 1988, *supra*). Targeted peptides, e.g., hormones or neurotransmitters, and proteins such as antibodies, or non-peptide molecules could be coupled to liposomes chemically.

[0073] Other molecules are also useful for facilitating transfection of a nucleic acid *in vivo*, such as a cationic oligopeptide (e.g., WO95/21931), peptides derived from DNA binding proteins (e.g., WO96/25508), or a cationic polymer (e.g., WO95/21931).

[0074] It is also possible to introduce a vector *in vivo* as a naked DNA plasmid (see U.S. Patents 5,693,622, 5,589,466 and 5,580,859). Receptor-mediated DNA delivery approaches can also be used (Curiel et al., 1992, Hum. Gene Ther. 3:147-154; and Wu and Wu, 1987, J. Biol. Chem. 262:4429-4432).

[0075] The term "transfection" means the uptake of exogenous or heterologous RNA or DNA by a cell. A cell has been

"transfected" by exogenous or heterologous RNA or DNA when such RNA or DNA has been introduced inside the cell. A cell has been "transformed" by exogenous or heterologous RNA or DNA when the transfected RNA or DNA effects a phenotypic change. The transforming RNA or DNA can be integrated (covalently linked) into chromosomal DNA making up the genome of the cell.

5 [0076] "Transformation" refers to the transfer of a nucleic acid fragment into the genome of a host organism, resulting in genetically stable inheritance. Host organisms containing the transformed nucleic acid fragments are referred to as "transgenic" or "recombinant" or "transformed" organisms.

[0077] The term "genetic region" will refer to a region of a nucleic acid molecule or a nucleotide sequence that comprises a gene encoding a polypeptide.

10 [0078] In addition, the recombinant vector comprising a polynucleotide according to the invention may include one or more origins for replication in the cellular hosts in which their amplification or their expression is sought, markers or selectable markers.

15 [0079] The term "selectable marker" means an identifying factor, usually an antibiotic or chemical resistance gene, that is able to be selected for based upon the marker gene's effect, i.e., resistance to an antibiotic, resistance to a herbicide, colorimetric markers, enzymes, fluorescent markers, and the like, wherein the effect is used to track the inheritance of a nucleic acid of interest and/or to identify a cell or organism that has inherited the nucleic acid of interest. Examples of selectable marker genes known and used in the art include: genes providing resistance to ampicillin, streptomycin, gentamycin, kanamycin, hygromycin, bialaphos herbicide, sulfonamide, and the like; and genes that are used as phenotypic markers, i.e., anthocyanin regulatory genes, isopentyl transferase gene, and the like.

20 [0080] The term "reporter gene" means a nucleic acid encoding an identifying factor that is able to be identified based upon the reporter gene's effect, wherein the effect is used to track the inheritance of a nucleic acid of interest, to identify a cell or organism that has inherited the nucleic acid of interest, and/or to measure gene expression induction or transcription. Examples of reporter genes known and used in the art include: luciferase (Luc), green fluorescent protein (GFP), chloramphenicol acetyltransferase (CAT),  $\beta$ -galactosidase (LacZ),  $\beta$ -glucuronidase (Gus), and the like. Selectable marker genes may also be considered reporter genes.

25 [0081] "Promoter" refers to a DNA sequence capable of controlling the expression of a coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic DNA segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental or physiological conditions. Promoters that cause a gene to be expressed in most cell types at most times are commonly referred to as "constitutive promoters". Promoters that cause a gene to be expressed in a specific cell type are commonly referred to as "cell-specific promoters" or "tissue-specific promoters". Promoters that cause a gene to be expressed at a specific stage of development or cell differentiation are commonly referred to as "developmentally-specific promoters" or "cell differentiation-specific promoters". Promoters that are induced and cause a gene to be expressed following exposure or treatment of the cell with an agent, biological molecule, chemical, ligand, light, or the like that induces the promoter are commonly referred to as "inducible promoters" or "regulatable promoters". It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, DNA fragments of different lengths may have identical promoter activity.

30 [0082] A "promoter sequence" is a DNA regulatory region capable of binding RNA polymerase in a cell and initiating transcription of a downstream (3' direction) coding sequence. For purposes of defining the present invention, the promoter sequence is bounded at its 3' terminus by the transcription initiation site and extends upstream (5' direction) to include the minimum number of bases or elements necessary to initiate transcription at levels detectable above background. Within the promoter sequence will be found a transcription initiation site (conveniently defined for example, by mapping with nuclelease S1), as well as protein binding domains (consensus sequences) responsible for the binding of RNA polymerase.

35 [0083] A coding sequence is "under the control" of transcriptional and translational control sequences in a cell when RNA polymerase transcribes the coding sequence into mRNA, which is then trans-RNA spliced (if the coding sequence contains introns) and translated into the protein encoded by the coding sequence.

40 [0084] "Transcriptional and translational control sequences" are DNA regulatory sequences, such as promoters, enhancers, terminators, and the like, that provide for the expression of a coding sequence in a host cell. In eukaryotic cells, polyadenylation signals are control sequences.

45 [0085] The term "response element" means one or more cis-acting DNA elements which confer responsiveness on a promoter mediated through interaction with the DNA-binding domains of the first chimeric gene. This DNA element may be either palindromic (perfect or imperfect) in its sequence or composed of sequence motifs or half sites separated by a variable number of nucleotides. The half sites can be similar or identical and arranged as either direct or inverted repeats or as a single half site or multimers of adjacent half sites in tandem. The response element may comprise a minimal promoter isolated from different organisms depending upon the nature of the cell or organism into which the

response element will be incorporated. The DNA binding domain of the first hybrid protein binds, in the presence or absence of a ligand, to the DNA sequence of a response element to initiate or suppress transcription of downstream gene(s) under the regulation of this response element. Examples of DNA sequences for response elements of the natural ecdysone receptor include: RRGG/TTCANTGAC/ACYY (see Cherbas L., et al., (1991), *Genes Dev.* 5, 120-131); AG-GTCAN<sub>n</sub>AGTCA, where N<sub>n</sub> can be one or more spacer nucleotides (see D'Avino PP., et al., (1995), *Mol. Cell. Endocrinol.*, 113,1-9); and GGGTTGAATGAATT (see Antoniewski C., et al., (1994), *Mol. Cell Biol.* 14, 4465-4474).

[0086] The term "operably linked" refers to the association of nucleic acid sequences on a single nucleic acid fragment so that the function of one is affected by the other. For example, a promoter is operably linked with a coding sequence when it is capable of affecting the expression of that coding sequence (i.e., that the coding sequence is under the transcriptional control of the promoter). Coding sequences can be operably linked to regulatory sequences in sense or antisense orientation.

[0087] The term "expression", as used herein, refers to the transcription and stable accumulation of sense (mRNA) or antisense RNA derived from a nucleic acid or polynucleotide. Expression may also refer to translation of mRNA into a protein or polypeptide.

[0088] The terms "cassette", "expression cassette" and "gene expression cassette" refer to a segment of DNA that can be inserted into a nucleic acid or polynucleotide at specific restriction sites or by homologous recombination. The segment of DNA comprises a polynucleotide that encodes a polypeptide of interest, and the cassette and restriction sites are designed to ensure insertion of the cassette in the proper reading frame for transcription and translation. "Transformation cassette" refers to a specific vector comprising a polynucleotide that encodes a polypeptide of interest and having elements in addition to the polynucleotide that facilitate transformation of a particular host cell. Casettes, expression cassettes, gene expression cassettes and transformation cassettes of the invention may also comprise elements that allow for enhanced expression of a polynucleotide encoding a polypeptide of interest in a host cell. These elements may include, but are not limited to: a promoter, a minimal promoter, an enhancer, a response element, a terminator sequence, a polyadenylation sequence, and the like.

[0089] For purposes of this invention, the term "gene switch" refers to the combination of a response element associated with a promoter, and an ECR based system which, in the presence of one or more ligands, modulates the expression of a gene into which the response element and promoter are incorporated.

[0090] The terms "modulate" and "modulates" mean to induce, reduce or inhibit nucleic acid or gene expression, resulting in the respective induction, reduction or inhibition of protein or polypeptide production.

[0091] The plasmids or vectors according to the invention may further comprise at least one promoter suitable for driving expression of a gene in a host cell. The term "expression vector" means a vector, plasmid or vehicle designed to enable the expression of an inserted nucleic acid sequence following transformation into the host. The cloned gene, i.e., the inserted nucleic acid sequence, is usually placed under the control of control elements such as a promoter, a minimal promoter, an enhancer, or the like. Initiation control regions or promoters, which are useful to drive expression of a nucleic acid in the desired host cell are numerous and familiar to those skilled in the art. Virtually any promoter capable of driving these genes is suitable for the present invention including but not limited to: viral promoters, plant promoters, bacterial promoters, animal promoters, mammalian promoters, synthetic promoters, constitutive promoters, tissue specific promoter, developmental specific promoters, inducible promoters, light regulated promoters; CYC1, HIS3, GAL1, GAL4, GAL10, ADHI, PGK, PHOS, GAPDH, ADCT, TRP1, URA3, LEU2, ENO, TPI, alkaline phosphatase promoters (useful for expression in *Saccharomyces*); AOX1 promoter (useful for expression in *Pichia*); b-lactamase, lac, ara, tet, trp, IP<sub>L</sub>, IP<sub>B</sub>, T7, tac, and trc promoters (useful for expression in *Escherichia coli*); and light regulated-, seed specific-, pollen specific-, ovary specific-, pathogenesis or disease related-, cauliflower mosaic virus 35S, CMV 35S minimal, cassava vein mosaic virus (CsMV), chlorophyll a/b binding protein, ribulose 1, 5-bisphosphate carboxylase, shoot-specific, root specific, chitinase, stress inducible, rice tungro bacilliform virus, plant super-promoter, potato leucine aminopeptidase, nitrate reductase, mannosamine synthase, nopaline synthase, ubiquitin, zein protein, and an-thocyanin promoters (useful for expression in plant cells); animal and mammalian promoters known in the art include, but are not limited to, the SV40 early (SV40e) promoter region, the promoter contained in the 3' long terminal repeat (LTR) of Rous sarcoma virus (RSV), the promoters of the E1A or major late promoter (MLP) genes of adenoviruses, the cytomegalovirus early promoter, the herpes simplex virus (HSV) thymidine kinase (TK) promoter, an elongation factor 1 alpha (EF1) promoter, a phosphoglycerate kinase (PGK) promoter, a ubiquitin (Ubc) promoter, an albumin promoter, the regulatory sequences of the mouse metallothionein-L promoter, and transcriptional control regions, the ubiquitous promoters (Hprt, vimentin,  $\alpha$ -actin, tubulin and the like), the promoters of the intermediate filaments (desmin, neurofilaments, keratin, GFAP, and the like), the promoters of therapeutic genes (of the MDR, CFTR or factor VIII type, and the like), and promoters that exhibit tissue specificity and have been utilized in transgenic animals, such as the elastase I gene control region which is active in pancreatic acinar cells; insulin gene control region active in pancreatic beta cells, immunoglobulin gene control region active in lymphoid cells, mouse mammary tumor virus control region active in testicular, breast, lymphoid and mast cells; albumin gene, Apo AI and Apo All control regions active in liver, alpha-fetoprotein gene control region active in liver, alpha 1-antitrypsin gene control region active in the liver, beta-globin gene

control region active in myeloid cells, myelin basic protein gene control region ) active in oligodendrocyte cells in the brain, myosin light chain-2 gene control region active in skeletal muscle, and gonadotropic releasing hormone gene control region active in the hypothalamus, pyruvate kinase promoter, villin promoter, promoter of the fatty acid binding intestinal protein, promoter of the smooth muscle cell  $\alpha$ -actin, and the like. In a preferred embodiment of the invention, the promoter is selected from the group consisting of a cauliflower mosaic virus 35S promoter, a cassava vein mosaic virus promoter, and a cauliflower mosaic virus 35S minimal promoter, an elongation factor 1 alpha (EF1) promoter, a phosphoglycerate kinase (PGK) promoter, a ubiquitin (Ubc) promoter, and an albumin promoter. In addition, these expression sequences may be modified by addition of enhancer or regulatory sequences and the like.

[0092] Enhancers that may be used in embodiments of the invention include but are not limited to: tobacco mosaic virus enhancer, cauliflower mosaic virus 35S enhancer, tobacco etch virus enhancer, ribulose 1, 5-bisphosphate carboxylase enhancer, rice tungro bacilliform virus enhancer, and other plant and viral gene enhancers, and the like.

[0093] Termination control regions, i.e., terminator or polyadenylation sequences, may also be derived from various genes native to the preferred hosts. Optionally, a termination site may be unnecessary, however, it is most preferred if included. In a preferred embodiment of the invention, the termination control region may be comprise or be derived from a synthetic sequence, synthetic polyadenylation signal, an SV40 late polyadenylation signal, an SV40 polyadenylation signal, a bovine growth hormone (BGH) polyadenylation signal, nopaline synthase (nos), cauliflower mosaic virus (CaMV), octopine synthase (ocs), Agrobacterium, viral, and plant terminator sequences, or the like.

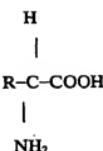
[0094] The terms "3' non-coding sequences" or "3' untranslated region (UTR)" refer to DNA sequences located downstream (3') of a coding sequence and may comprise polyadenylation [poly(A)] recognition sequences and other sequences encoding regulatory signals capable of affecting mRNA processing or gene expression. The polyadenylation signal is usually characterized by affecting the addition of polyadenylic acid tracts to the 3' end of the mRNA precursor.

[0095] "Regulatory region" means a nucleic acid sequence which regulates the expression of a second nucleic acid sequence. A regulatory region may include sequences which are naturally responsible for expressing a particular nucleic acid (a homologous region) or may include sequences of a different origin that are responsible for expressing different proteins or even synthetic proteins (a heterologous region). In particular, the sequences can be sequences of prokaryotic, eukaryotic, or viral genes or derived sequences that stimulate or repress transcription of a gene in a specific or non-specific manner and in an inducible or non-inducible manner. Regulatory regions include origins of replication, RNA splice sites, promoters, enhancers, transcriptional termination sequences, and signal sequences which direct the polypeptide into the secretory pathways of the target cell.

[0096] A regulatory region from a "heterologous source" is a regulatory region that is not naturally associated with the expressed nucleic acid. Included among the heterologous regulatory regions are regulatory regions from a different species, regulatory regions from a different gene, hybrid regulatory sequences, and regulatory sequences which do not occur in nature, but which are designed by one having ordinary skill in the art.

[0097] "RNA transcript" refers to the product resulting from RNA polymerase-catalyzed transcription of a DNA sequence. When the RNA transcript is a perfect complementary copy of the DNA sequence, it is referred to as the primary transcript or it may be a RNA sequence derived from post-transcriptional processing of the primary transcript and is referred to as the mature RNA. "Messenger RNA (mRNA)" refers to the RNA that is without introns and that can be translated into protein by the cell. "cDNA" refers to a double-stranded DNA that is complementary to and derived from mRNA. "Sense" RNA refers to RNA transcript that includes the mRNA and so can be translated into protein by the cell. "Antisense RNA" refers to a RNA transcript that is complementary to all or part of a target primary transcript or mRNA and that blocks the expression of a target gene. The complementarity of an antisense RNA may be with any part of the specific gene transcript, i.e., at the 5' non-coding sequence, 3' non-coding sequence, or the coding sequence. "Functional RNA" refers to antisense RNA, ribozyme RNA, or other RNA that is not translated yet has an effect on cellular processes.

[0098] A "polypeptide" is a polymeric compound comprised of covalently linked amino acid residues. Amino acids have the following general structure:



Amino acids are classified into seven groups on the basis of the side chain R: (1) aliphatic side chains, (2) side chains containing a hydroxyl (OH) group, (3) side chains containing sulfur atoms, (4) side chains containing an acidic or amide

5

10

15

20

25

30

35

40

45

50

55

group, (5) side chains containing a basic group, (6) side chains containing an aromatic ring, and (7) proline, an imino acid in which the side chain is fused to the amino group. A polypeptide of the invention preferably comprises at least about 14 amino acids.

[0099] A "protein" is a polypeptide that performs a structural or functional role in a living cell.

[0100] An "isolated polypeptide" or "isolated protein" is a polypeptide or protein that is substantially free of those compounds that are normally associated therewith in its natural state (e.g., other proteins or polypeptides, nucleic acids, carbohydrates, lipids). "Isolated" is not meant to exclude artificial or synthetic mixtures with other compounds, or the presence of impurities which do not interfere with biological activity, and which may be present, for example, due to incomplete purification, addition of stabilizers, or compounding into a pharmaceutically acceptable preparation.

[0101] "Fragment" of a polypeptide according to the invention will be understood to mean a polypeptide whose amino acid sequence is shorter than that of the reference polypeptide and which comprises, over the entire portion with these reference polypeptides, an identical amino acid sequence. Such fragments may, where appropriate, be included in a larger polypeptide of which they are a part. Such fragments of a polypeptide according to the invention may have a length of 10, 15, 20, 30 to 40, 50, 100, 200 or 300 amino acids.

[0102] A "variant" of a polypeptide or protein is any analogue, fragment, derivative, or mutant which is derived from a polypeptide or protein and which retains at least one biological property of the polypeptide or protein. Different variants of the polypeptide or protein may exist in nature. These variants may be allelic variations characterized by differences in the nucleotide sequences of the structural gene coding for the protein, or may involve differential splicing or post-translational modification. The skilled artisan can produce variants having single or multiple amino acid substitutions, deletions, additions, or replacements. These variants may include, *inter alia*: (a) variants in which one or more amino acid residues are substituted with conservative or non-conservative amino acids, (b) variants in which one or more amino acids are added to the polypeptide or protein, (c) variants in which one or more of the amino acids includes a substituent group, and (d) variants in which the polypeptide or protein is fused with another polypeptide such as serum albumin. The techniques for obtaining these variants, including genetic (suppressions, deletions, mutations, etc.), chemical, and enzymatic techniques, are known to persons having ordinary skill in the art. A variant polypeptide preferably comprises at least about 14 amino acids.

[0103] A "heterologous protein" refers to a protein not naturally produced in the cell.

[0104] A "mature protein" refers to a post-translationally processed polypeptide; i.e., one from which any pre- or propeptides present in the primary translation product have been removed. "Precursor" protein refers to the primary product of translation of mRNA; i.e., with pre- and propeptides still present Pre- and propeptides may be but are not limited to intracellular localization signals.

[0105] The term "signal peptide" refers to an amino terminal polypeptide preceding the secreted mature protein. The signal peptide is cleaved from and is therefore not present in the mature protein. Signal peptides have the function of directing and translocating secreted proteins across cell membranes. Signal peptide is also referred to as signal protein.

[0106] A "signal sequence" is included at the beginning of the coding sequence of a protein to be expressed on the surface of a cell. This sequence encodes a signal peptide, N-terminal to the mature polypeptide, that directs the host cell to translocate the polypeptide. The term "translocation signal sequence" is used herein to refer to this sort of signal sequence. Translocation signal sequences can be found associated with a variety of proteins native to eukaryotes and prokaryotes, and are often functional in both types of organisms.

[0107] The term "homology" refers to the percent of identity between two polynucleotide or two polypeptide moieties. The correspondence between the sequence from one moiety to another can be determined by techniques known to the art. For example, homology can be determined by a direct comparison of the sequence information between two polypeptide molecules by aligning the sequence information and using readily available computer programs. Alternatively, homology can be determined by hybridization of polynucleotides under conditions that form stable duplexes between homologous regions, followed by digestion with single-stranded-specific nuclease(s) and size determination of the digested fragments.

[0108] As used herein, the term "homologous" in all its grammatical forms and spelling variations refers to the relationship between proteins that possess a "common evolutionary origin," including proteins from superfamilies (e.g., the immunoglobulin superfamily) and homologous proteins from different species (e.g., myosin light chain, etc.) (Reeck et al., 1987, Cell 50:667.). Such proteins (and their encoding genes) have sequence homology, as reflected by their high degree of sequence similarity.

[0109] Accordingly, the term "sequence similarity" in all its grammatical forms refers to the degree of identity or correspondence between nucleic acid or amino acid sequences of proteins that may or may not share a common evolutionary origin (see Reeck et al., 1987, Cell 50:667.). As used herein, the term "homologous" in all its grammatical forms and spelling variations refers to the relationship between proteins that possess a "common evolutionary origin," including proteins from superfamilies and homologous proteins from different species (Reeck et al., *supra*). Such proteins (and their encoding genes) have sequence homology, as reflected by their high degree of sequence similarity. However, in common usage and in the instant application, the term "homologous," when modified with an adverb such as "highly,"

may refer to sequence similarity and not a common evolutionary origin.

[0110] In a specific embodiment, two DNA sequences are "substantially homologous" or "substantially similar" when at least about 50% (preferably at least about 75%, and most preferably at least about 90 or 95%) of the nucleotides match over the defined length of the DNA sequences. Sequences that are substantially homologous can be identified by comparing the sequences using standard software available in sequence data banks, or in a Southern hybridization experiment under, for example, stringent conditions as defined for that particular system. Defining appropriate hybridization conditions is within the skill of the art. See, e.g., Sambrook *et al.*, 1989, *supra*.

[0111] As used herein, "substantially similar" refers to nucleic acid fragments wherein changes in one or more nucleotide bases results in substitution of one or more amino acids, but do not affect the functional properties of the protein encoded by the DNA sequence. "Substantially similar" also refers to nucleic acid fragments wherein changes in one or more nucleotide bases does not affect the ability of the nucleic acid fragment to mediate alteration of gene expression by antisense or co-suppression technology. "Substantially similar" also refers to modifications of the nucleic acid fragments of the instant invention such as deletion or insertion of one or more nucleotide bases that do not substantially affect the functional properties of the resulting transcript. It is therefore understood that the invention encompasses more than the specific exemplary sequences. Each of the proposed modifications is well within the routine skill in the art, as is determination of retention of biological activity of the encoded products.

[0112] Moreover, the skilled artisan recognizes that substantially similar sequences encompassed by this invention are also defined by their ability to hybridize, under stringent conditions (0.1X SSC, 0.1% SDS, 65°C and washed with 2X SSC, 0.1% SDS followed by 0.1X SSC, 0.1% SDS), with the sequences exemplified herein. Substantially similar nucleic acid fragments of the instant invention are those nucleic acid fragments whose DNA sequences are at least 70% identical to the DNA sequence of the nucleic acid fragments reported herein. Preferred substantially nucleic acid fragments of the instant invention are those nucleic acid fragments whose DNA sequences are at least 80% identical to the DNA sequence of the nucleic acid fragments reported herein. More preferred nucleic acid fragments are at least 90% identical to the DNA sequence of the nucleic acid fragments reported herein. Even more preferred are nucleic acid fragments that are at least 95% identical to the DNA sequence of the nucleic acid fragments reported herein.

[0113] Two amino acid sequences are "substantially homologous" or "substantially similar" when greater than about 40% of the amino acids are identical, or greater than 60% are similar (functionally identical). Preferably, the similar or homologous sequences are identified by alignment using, for example, the GCG (Genetics Computer Group, Program Manual for the GCG Package, Version 7, Madison, Wisconsin) pileup program.

[0114] The term "corresponding to" is used herein to refer to similar or homologous sequences, whether the exact position is identical or different from the molecule to which the similarity or homology is measured. A nucleic acid or amino acid sequence alignment may include spaces. Thus, the term "corresponding to" refers to the sequence similarity, and not the numbering of the amino acid residues or nucleotide bases.

[0115] A "substantial portion" of an amino acid or nucleotide sequence comprises enough of the amino acid sequence of a polypeptide or the nucleotide sequence of a gene to putatively identify that polypeptide or gene, either by manual evaluation of the sequence by one skilled in the art, or by computer-automated sequence comparison and identification using algorithms such as BLAST (Basic Local Alignment Search Tool; Altschul, S. F., *et al.*, (1993) *J. Mol. Biol.* 215: 403-410; see also [www.ncbi.nlm.nih.gov/BLAST/](http://www.ncbi.nlm.nih.gov/BLAST/)). In general, a sequence of ten or more contiguous amino acids or thirty or more nucleotides is necessary in order to putatively identify a polypeptide or nucleic acid sequence as homologous to a known protein or gene. Moreover, with respect to nucleotide sequences, gene specific oligonucleotide probes comprising 20-30 contiguous nucleotides may be used in sequence-dependent methods of gene identification (e.g., Southern hybridization) and isolation (e.g., *in situ* hybridization of bacterial colonies or bacteriophage plaques). In addition, short oligonucleotides of 12-15 bases may be used as amplification primers in PCR in order to obtain a particular nucleic acid fragment comprising the primers. Accordingly, a "substantial portion" of a nucleotide sequence comprises enough of the sequence to specifically identify and/or isolate a nucleic acid fragment comprising the sequence.

[0116] The term "percent identity", as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be, as determined by the match between strings of such sequences. "Identity" and "similarity" can be readily calculated by known methods, including but not limited to those described in: *Computational Molecular Biology* (Lesk, A. M., ed.) Oxford University Press, New York (1988); *Biocomputing: Informatics and Genome Projects* (Smith, D. W., ed.) Academic Press, New York (1993); *Computer Analysis of Sequence Data, Part I* (Griffin, A. M., and Griffin, H. G., eds.) Humana Press, New Jersey (1994); *Sequence Analysis in Molecular Biology* (von Heijne, G., ed.) Academic Press (1987); and *Sequence Analysis Primer* (Gribkov, M. and Devereux, J., eds.) Stockton Press, New York (1991). Preferred methods to determine identity are designed to give the best match between the sequences tested. Methods to determine identity and similarity are codified in publicly available computer programs. Sequence alignments and percent identity calculations may be performed using the Megalign program of the LASERGENE bioinformatics computing suite (DNASTAR Inc., Madison, WI). Multiple alignment of the sequences may be performed using the Clustal method of alignment (Higgins

and Sharp (1989) CABIOS. 5:151-153) with the default parameters (GAP PENALTY=10, GAP LENGTH PENALTY=10). Default parameters for pairwise alignments using the Clustal method may be selected: KTUPLE 1, GAP PENALTY=3, WINDOW=5 and DIAGONALS SAVED=5.

**[0117]** The term "sequence analysis software" refers to any computer algorithm or software program that is useful for the analysis of nucleotide or amino acid sequences. "Sequence analysis software" may be commercially available or independently developed. Typical sequence analysis software will include but is not limited to the GCG suite of programs (Wisconsin Package Version 9.0, Genetics Computer Group (GCG), Madison, WI), BLASTP, BLASTN, BLASTX (Altschul et al., *J. Mol. Biol.* 215:403-410 (1990), and DNASTAR (DNASTAR, Inc. 1228 S. Park St, Madison, WI 53715 USA). Within the context of this application it will be understood that where sequence analysis software is used for analysis, that the results of the analysis will be based on the "default values" of the program referenced, unless otherwise specified. As used herein "default values" will mean any set of values or parameters which originally load with the software when first initialized.

**[0118]** "Synthetic genes" can be assembled from oligonucleotide building blocks that are chemically synthesized using procedures known to those skilled in the art. These building blocks are ligated and annealed to form gene segments that are then enzymatically assembled to construct the entire gene. "Chemically synthesized", as related to a sequence of DNA, means that the component nucleotides were assembled *in vitro*. Manual chemical synthesis of DNA may be accomplished using well established procedures, or automated chemical synthesis can be performed using one of a number of commercially available machines. Accordingly, the genes can be tailored for optimal gene expression based on optimization of nucleotide sequence to reflect the codon bias of the host cell. The skilled artisan appreciates the likelihood of successful gene expression if codon usage is biased towards those codons favored by the host. Determination of preferred codons can be based on a survey of genes derived from the host cell where sequence information is available.

#### GENE EXPRESSION MODULATION SYSTEM OF THE INVENTION

**[0119]** Applicants have now shown that separating the transactivation and DNA binding domains by placing them on two different proteins results in greatly reduced background activity in the absence of a ligand and significantly increased activity over background in the presence of a ligand. Applicants' improved gene expression system comprises two chimeric gene expression; the first encoding a DNA binding domain fused to a nuclear receptor polypeptide and the second encoding a transactivation domain fused to a nuclear receptor polypeptide. The interaction of the first protein with the second protein effectively tethers the DNA binding domain to the transactivation domain. Since the DNA binding and transactivation domains reside on two different molecules, the background activity in the absence of ligand is greatly reduced.

**[0120]** In general, the inducible gene expression modulation system of the invention comprises a) a first chimeric gene that is capable of being expressed in a host cell comprising a polynucleotide sequence that encodes a first hybrid polypeptide comprising i) a DNA binding domain that recognizes a response element associated with a gene whose expression is to be modulated; and ii) a ligand binding domain comprising the ligand binding domain from a nuclear receptor, and b) a second chimeric gene that is capable of being expressed in the host cell comprising a polynucleotide sequence that encodes a second hybrid polypeptide comprising: i) a transactivation domain; and ii) a ligand binding domain comprising the ligand binding domain from a nuclear receptor other than ultraspiracle (USP); wherein the transactivation domain are from other than EcR, RXR, or USP; and wherein the ligand binding domains from the first hybrid polypeptide and the second hybrid polypeptide are different and dimerize.

**[0121]** This two-hybrid system exploits the ability of a pair of interacting proteins to bring the transcription activation domain into a more favorable position relative to the DNA binding domain such that when the DNA binding domain binds to the DNA binding site on the gene, the transactivation domain more effectively activates the promoter (see, for example, U.S. Patent No. 5,283,173). This two-hybrid system is a significantly improved inducible gene expression modulation system compared to the two systems disclosed in International Patent Applications PCT/US97/05330 and PCT/US98/14215.

**[0122]** The ecdysone receptor-based gene expression modulation system of the invention may be either heterodimeric and homodimeric. A functional EcR complex generally refers to a heterodimeric protein complex consisting of two members of the steroid receptor family, an ecdysone receptor protein obtained from various insects, and an ultraspiracle (USP) protein or the vertebrate homolog of USP, retinoid X receptor protein (see Yao, et al. (1993) *Nature* 366, 476-479; Yao, et al., (1992) *Cell* 71, 63-72). However, the complex may also be a homodimer as detailed below. The functional ecdysteroid receptor complex may also include additional protein(s) such as immunophilins. Additional members of the steroid receptor family of proteins, known as transcriptional factors (such as DHR38 or *betaFTZ-1*), may also be ligand dependent or independent partners for EcR, USP, and/or RXR. Additionally, other cofactors may be required such as proteins generally known as coactivators (also termed adapters or mediators). These proteins do not bind sequence-specifically to DNA and are not involved in basal transcription. They may exert their effect on transcription activation

through various mechanisms, including stimulation of DNA-binding of activators, by affecting chromatin structure, or by mediating activator-initiation complex interactions. Examples of such coactivators include RIP140, TIF1, RAP46Bag-1, ARA70, SRC-1/NCoA-1, TIF2/GRIP/NCoA-2, ACTR/AIB 1/RAC3/pCIP as well as the promiscuous co activator C response element B binding protein, CBP/p300 (for review see Glass et al. *Curr. Opin. Cell Biol.* 9: 222-232, 1997). Also, protein cofactors generally known as corepressors (also known as repressors, silencers, or silencing mediators) may be required to effectively inhibit transcriptional activation in the absence of ligand. These corepressors may interact with the unliganded ecdysone receptor to silence the activity at the response element. Current evidence suggests that binding of ligand changes the conformation of the receptor, which results in release of the corepressor and recruitment of the above described coactivators, thereby, abolishing their silencing activity. Examples of corepressors include N-CoR and SMRT (for review, see Horwitz et al. *Mol Endocrinol.* 10: 1167-1177, 1996). These cofactors may either be endogenous within the cell or organism, or may be added exogenously as transgenes to be expressed in either a regulated or unregulated fashion. Homodimer complexes of the ecdysone receptor protein, USP, or RXR may also be functional under some circumstances.

**[0123]** The ecdysone receptor complex typically includes proteins which are members of the nuclear receptor superfamily wherein all members are characterized by the presence of an amino-terminal transactivation domain, a DNA binding domain ("DBD"), and a ligand binding domain ("LBD") separated from the DBD by a hinge region. As used herein, the term "DNA binding domain" comprises a minimal peptide sequence of a DNA binding protein, up to the entire length of a DNA binding protein, so long as the DNA binding domain functions to associate with a particular response element. Members of the nuclear receptor superfamily are also characterized by the presence of four or five domains: A/B, C, D, E, and in some members F (see Evans, *Science* 240:889-895 (1988)). The "AB" domain corresponds to the transactivation domain, "C" corresponds to the DNA binding domain, "D" corresponds to the hinge region, and "E" corresponds to the ligand binding domain. Some members of the family may also have another transactivation domain on the carboxy-terminal side of the LBD corresponding to "F".

**[0124]** The DBD is characterized by the presence of two cysteine zinc fingers between which are two amino acid motifs, the P-box and the D-box, which confer specificity for ecdysone response elements. These domains may be either native, modified, or chimeras of different domains of heterologous receptor proteins. This EcR receptor, like a subset of the steroid receptor family, also possesses less well defined regions responsible for heterodimerization properties. Because the domains of EcR, USP, and RXR are modular in nature, the LBD, DBD, and transactivation domains may be interchanged.

**[0125]** Gene switch systems are known that incorporate components from the ecdysone receptor complex. However, in these known systems, whenever EcR is used it is associated with native or modified DNA binding domains and transactivation domains on the same molecule. USP or RXR are typically used as silent partners. We have now shown that when DNA binding domains and transactivation domains are on the same molecule the background activity in the absence of ligand is high and that such activity is dramatically reduced when DNA binding domains and transactivation domains are on different molecules, that is, on each of two partners of a heterodimeric or homodimeric complex. This two-hybrid system also provides improved sensitivity to non-steroidal ligands for example, diacylhydrazines, when compared to steroidal ligands for example, ponasterone A ("PonA") or muristerone A ("MurA"). That is, when compared to steroids, the non-steroidal ligands provide higher activity at a lower concentration. In addition, since transactivation based on EcR gene switches is often cell-line dependent, it is easier to tailor switching system to obtain maximum transactivation capability for each application. Furthermore, this two-hybrid system avoids some side effects due to overexpression of RXR that often occur when unmodified RXR is used as a switching partner. In this two-hybrid system, native DNA binding and transactivation domains of EcR or RXR are eliminated. As a result, these chimeric molecules have less chance of interacting with other steroid hormone receptors present in the cell resulting in reduced side effects.

**[0126]** Specifically, Applicants' invention relates to a gene expression modulation system comprising: a) a first gene expression cassette that is capable of being expressed in a host cell, wherein the first gene expression cassette comprises a polynucleotide that encodes a first polypeptide comprising i) a DNA-binding domain that recognizes a response element associated with a gene whose expression is to be modulated; and ii) a ligand binding domain comprising a ligand binding domain from a nuclear receptor; and b) a second gene expression cassette that is capable of being expressed in the host cell, wherein the second gene expression cassette comprises a polynucleotide sequence that encodes a second polypeptide comprising i) a transactivation domain; and ii) a ligand binding domain comprising a ligand binding domain from a nuclear receptor other than ultraspiracle (USP); wherein the DNA binding domain and the transactivation domain are from other than EcR, RXR, or USP; wherein the ligand binding domains from the first polypeptide and the second polypeptide are different and dimerize.

**[0127]** The present invention also relates to a gene expression modulation system according to the present invention further comprising c) a third gene expression cassette comprising: i) the response element to which the DNA-binding domain of the first polypeptide binds; ii) a promoter that is activated by the transactivation domain of the second polypeptide; and iii) the gene whose expression is to be modulated.

**[0128]** In a specific embodiment, the gene whose expression is to be modulated is a homologous gene with respect

to the host cell. In another specific embodiment, the gene whose expression is to be modulated is a heterologous gene with respect to the host cell.

[0129] In a specific embodiment, the ligand binding domain of the first polypeptide comprises an ecdysone receptor ligand binding domain.

5 [0130] In another specific embodiment, the ligand binding domain of the first polypeptide comprises a retinoid X receptor ligand binding domain.

[0131] In a specific embodiment, the ligand binding domain of the second polypeptide comprises an ecdysone receptor ligand binding domain.

10 [0132] In another specific embodiment, the ligand binding domain of the second polypeptide comprises a retinoid X receptor ligand binding domain.

[0133] In a preferred embodiment, the ligand binding domain of the first polypeptide comprises an ecdysone receptor ligand binding domain, and the ligand binding domain of the second polypeptide comprises a retinoid X receptor ligand binding domain.

15 [0134] In another preferred embodiment, the ligand binding domain of the first polypeptide is from a retinoid X receptor polypeptide, and the ligand binding domain of the second polypeptide is from an ecdysone receptor polypeptide.

[0135] Preferably, the ligand binding domain is an EcR or RXR related steroid/thyroid hormone nuclear receptor family member ligand binding domain, or analogs, combinations, or modifications thereof. More preferably, the LBD is from EcR or RXR. Even more preferably, the LBD is from a truncated EcR or RXR. A truncation mutation may be made by any method used in the art, including but not limited to restriction endonuclease digestion/deletion, PCR-mediated oligonucleotide-directed deletion, chemical mutagenesis, UV strand breakage, and the like.

20 [0136] Preferably, the EcR is an insect EcR selected from the group consisting of a Lepidopteran EcR, a Dipteran EcR, an Arthropod EcR, a Homopteran EcR and a Hemipteran EcR. More preferably, the EcR for use is a spruce budworm *Choristoneura fumiferana* EcR ("CfEcR"), a *Tenebrio molitor* EcR ("TmEcR"), a *Manduca sexta* EcR ("MsEcR"), a *Heliothis virescens* EcR ("HvEcR"), a silk moth *Bombyx mori* EcR ("BmEcR"), a fruit fly *Drosophila melanogaster* EcR ("DmEcR"), a mosquito *Aedes aegypti* EcR ("AaEcR"), a blowfly *Lucilia capitata* EcR ("LcEcR"), a Mediterranean fruit fly *Ceratitis capitata* EcR ("CcEcR"), a locust *Locusta migratoria* EcR ("LmEcR"), an aphid *Myzus persicae* EcR ("MpEcR"), a fiddler crab *Uca pugilator* EcR ("UpEcR"), or an ixodid tick *Amblyomma americanum* EcR ("AaEcR"). Even more preferably, the LBD is from spruce budworm (*Choristoneura fumiferana*) EcR ("CfEcR") or fruit fly *Drosophila melanogaster* EcR ("DmEcR").

30 [0137] Preferably, the LBD is from a truncated insect EcR. The insect EcR polypeptide truncation comprises a deletion of at least 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, 150, 155, 160, 165, 170, 175, 180, 185, 190, 195, 200, 205, 210, 215, 220, 225, 230, 235, 240, 245, 250, 255, 260, or 265 amino acids. More preferably, the insect EcR polypeptide truncation comprises a deletion of at least a partial polypeptide domain. Even more preferably, the insect EcR polypeptide truncation comprises a deletion of at least an entire polypeptide domain. In a specific embodiment, the insect EcR polypeptide truncation comprises a deletion of at least an A/B-domain deletion, a C-domain deletion, a D-domain deletion, an E-domain deletion, an F-domain deletion, an A/B/C-domains deletion, an AB/I/2-C-domains deletion, an A/B/C/D-domains deletion, an AB/C/D/F-domains deletion, an A/B/F-domains, and an A/B/C/F-domains deletion. A combination of several complete and/or partial domain deletions may also be performed.

40 [0138] In a preferred embodiment, the ecdysone receptor ligand binding domain is encoded by a polynucleotide comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, and SEQ ID NO: 10.

[0139] In another preferred embodiment, the ecdysone receptor ligand binding domain comprises a polypeptide sequence selected from the group consisting of SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, and SEQ ID NO: 20.

45 [0140] Preferably, the RXR polypeptide is a mouse *Mus musculus* RXR ("MmRXR") or a human *Homo sapiens* RXR ("HsRXR"). The RXR polypeptide may be an RXR<sub>α</sub>, RXR<sub>β</sub>, or RXR<sub>isoform</sub>.

[0141] Preferably, the LBD is from a truncated RXR. The RXR polypeptide truncation comprises a deletion of at least 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135,

50 140, 145, 150, 155, 160, 165, 170, 175, 180, 185, 190, 195, 200, 205, 210, 215, 220, 225, 230, 235, 240, 245, 250, 255, 260, or 265 amino acids. More preferably, the RXR polypeptide truncation comprises a deletion of at least a partial polypeptide domain. Even more preferably, the RXR polypeptide truncation comprises a deletion of at least an entire polypeptide domain. In a specific embodiment, the RXR polypeptide truncation comprises a deletion of at least an A/B-domain deletion, a C-domain deletion, a D-domain deletion, an E-domain deletion, an F-domain - deletion, an A/B/C-domains deletion, an AB/I/2-C-domains deletion, an A/B/C/D-domains deletion, an A/B/C/D/F-domains deletion, an A/B/F-domains, and an A/B/C/F-domains deletion. A combination of several complete and/or partial domain deletions may also be performed.

55 [0142] In a preferred embodiment, the retinoid X receptor ligand binding domain is encoded by a polynucleotide

comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, and SEQ ID NO: 30.

**[0143]** In another preferred embodiment, the retinoid X receptor ligand binding domain comprises a polypeptide sequence selected from the group consisting of SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39, and SEQ ID NO: 40.

**[0144]** For purposes of this invention EcR and RXR also include synthetic and chimeric EcR and RXR and their homologs.

**[0145]** The DNA binding domain can be any DNA binding domain with a known response element, including synthetic and chimeric DNA binding domains, or analogs, combinations, or modifications thereof. Preferably, the DBD is a GAL4 DBD, a LexA DBD, a transcription factor DBD, a steroid/thyroid hormone nuclear receptor superfamily member DBD, a bacterial LacZ DBD, or a yeast put DBD. More preferably, the DBD is a GAL4 DBD [SEQ ID NO: 41 (polynucleotide) or SEQ ID NO: 42 (polypeptide)] or a LexA DBD [SEQ ID NO: 43 (polynucleotide) or SEQ ID NO: 44 (polypeptide)].

**[0146]** The transactivation domain (abbreviated "AD" or "TA") may be any steroid/thyroid hormone nuclear receptor AD, synthetic or chimeric AD, polyglutamyl AD, basic or acidic amino acid AD, a VP16 AD, a GAL4 AD, an NF-KB AD, a BP64 AD, or an analog, combination, or modification thereof. Preferably, the AD is a synthetic or chimeric AD, or is obtained from a VP16, GAL4, or NF-KB. Most preferably, the AD is a VP16 AD [SEQ ID NO: 45 (polynucleotide) or SEQ ID NO: 46 (polypeptide)].

**[0147]** The response element ("RE") may be any response element with a known DNA binding domain, or an analog, combination, or modification thereof. Preferably, the RE is an RE from GAL4 ("GAL4RE"), LexA, a steroid/thyroid hormone nuclear receptor RE, or a synthetic RE that recognizes a synthetic DNA binding domain. More preferably, the RE is a GAL4RE comprising a polynucleotide sequence of SEQ ID NO: 47 or a LexA 8X operon comprising a polynucleotide sequence of SEQ ID NO: 48. Preferably, the first hybrid protein is substantially free of a transactivation domain and the second hybrid protein is substantially free of a DNA binding domain. For purposes of this invention, "substantially free" means that the protein in question does not contain a sufficient sequence of the domain in question to provide activation or binding activity.

**[0148]** The ligands for use in the present invention as described below, when combined with the ligand binding domain of an EcR, USP, RXR, or another polypeptide which in turn are bound to the response element linked to a gene, provide the means for external temporal regulation of expression of the gene. The binding mechanism or the order in which the various components of this invention bind to each other, that is, ligand to receptor, first polypeptide to response element, second polypeptide to promoter, etc., is not critical. Binding of the ligand to the ligand binding domains of an EcR, USP, RXR, or another protein, enables expression or suppression of the gene. This mechanism does not exclude the potential for ligand binding to EcR, USP, or RXR, and the resulting formation of active homodimer complexes (e.g. EcR+EcR or USP+USP). Preferably, one or more of the receptor domains can be varied producing a chimeric gene switch. Typically, one or more of the three domains, DBD, LBD, and transactivation domain, may be chosen from a source different than the source of the other domains so that the chimeric genes and the resulting hybrid proteins are optimized in the chosen host cell or organism for transactivating activity, complementary binding of the ligand, and recognition of a specific response element. In addition, the response element itself can be modified or substituted with response elements for other DNA binding protein domains such as the GAL-4 protein from yeast (see Sadowski, et al. (1988) *Nature*, 335: 563-564) or LexA protein from *E. coli* (see Brent and Ptashne (1985), *Cell*, 43:729-736), or synthetic response elements specific for targeted interactions with proteins designed, modified, and selected for such specific interactions (see, for example, Kim, et al. (1997), *Proc. Natl. Acad Sci., USA*, 94:3616-3620) to accommodate chimeric receptors. Another advantage of chimeric systems is that they allow choice of a promoter used to drive the gene expression according to a desired end result. Such double control can be particularly important in areas of gene therapy, especially when cytotoxic proteins are produced, because both the timing of expression as well as the cells wherein expression occurs can be controlled. When genes, operatively linked to a suitable promoter, are introduced into the cells of the subject, expression of the exogenous genes is controlled by the presence of the system of this invention. Promoters may be constitutively or inducibly regulated or may be tissue-specific (that is, expressed only in a particular type of cells) or specific to certain developmental stages of the organism.

## 50 GENE EXPRESSION CASSETTES OF THE INVENTION

**[0149]** The novel ecdysone receptor-based inducible gene expression system of the invention comprises a novel gene expression cassette that is capable of being expressed in a host cell, wherein the gene expression cassette comprises a polynucleotide encoding a hybrid polypeptide. Thus, Applicants' invention also provides novel gene expression cassettes for use in the gene expression system of the invention.

**[0150]** Specifically, the present invention provides a gene expression cassette comprising a polynucleotide encoding a hybrid polypeptide. The hybrid polypeptide comprises either 1) a DNA-binding domain that recognizes a response element and a ligand binding domain of a nuclear receptor or 2) a transactivation domain and a ligand binding domain

of a nuclear receptor, wherein the transactivation domain is from a nuclear receptor other than an EcR, an RXR, or a USP. [0151] In a specific embodiment, the gene expression cassette encodes a hybrid polypeptide comprising a DNA-binding domain that recognizes a response element and an ecdysone receptor ligand binding domain, wherein the DNA binding domain is from a nuclear receptor other than an ecdysone receptor.

[0152] In another specific embodiment, the gene expression cassette encodes a hybrid polypeptide comprising a DNA binding domain that recognizes a response element and a retinoid X receptor ligand binding domain, wherein the DNA binding domain is from a nuclear receptor other than a retinoid X receptor.

[0153] The DNA binding domain can be any DNA binding domain with a known response element, including synthetic and chimeric DNA binding domains, or analogs, combinations, or modifications thereof. Preferably, the DBD is a GAL4 DBD, a LexA DBD, a transcription factor DBD, a steroid/thyroid hormone nuclear receptor superfamily member DBD, a bacterial LacZ DBD, or a yeast put DBD. More preferably, the DBD is a GAL4 DBD [SEQ ID NO: 41 (polynucleotide) or SEQ ID NO: 42 (polypeptide)] or a LexA DBD [SEQ ID NO: 43 (polynucleotide) or SEQ ID NO: 44 (polypeptide)].

[0154] In another specific embodiment, the gene expression cassette encodes a hybrid polypeptide comprising a transactivation domain and an ecdysone receptor ligand binding domain, wherein the transactivation domain is from a nuclear receptor other than an ecdysone receptor.

[0155] In another specific embodiment, the gene expression cassette encodes a hybrid polypeptide comprising a transactivation domain and a retinoid X receptor ligand binding domain, wherein the transactivation domain is from a nuclear receptor other than a retinoid X receptor.

[0156] The transactivation domain (abbreviated "AD" or "TA") may be any steroid/thyroid hormone nuclear receptor AD, synthetic or chimeric AD, polyglutamine AD, basic or acidic amino acid AD, a VP16 AD, a GAL4 AD, an NF- $\kappa$ B AD, a BP64 AD, or an analog, combination, or modification thereof. Preferably, the AD is a synthetic or chimeric AD, or is obtained from a VP16, GAL4, or NF- $\kappa$ B. Most preferably, the AD is a VP16 AD [SEQ ID NO: 45 (polynucleotide) or SEQ ID NO: 46 (polypeptide)].

[0157] Preferably, the ligand binding domain is an EcR or RXR related steroid/thyroid hormone nuclear receptor family member ligand binding domain, or analogs, combinations, or modifications thereof. More preferably, the LBD is from EcR or RXR. Even more preferably, the LBD is from a truncated EcR or RXR.

[0158] Preferably, the EcR is an insect EcR selected from the group consisting of a Lepidopteran EcR, a Dipteran EcR, an Arthropod EcR, a Homopteran EcR and a Hemipteran EcR. More preferably, the EcR for use is a spruce budworm *Choristoneura fumiferana* EcR ("CfEcR"), a *Tenebrio molitor* EcR ("TmEcR"), a *Manduca sexta* EcR ("MsEcR"), a *Heliothis virescens* EcR ("HvEcR"), a silk moth *Bombyx mori* EcR ("BmEcR"), a fruit fly *Drosophila melanogaster* EcR ("DmEcR"), a mosquito *Aedes aegypti* EcR ("AaEcR"), a blowfly *Lucilia capitata* EcR ("LcEcR"), a Mediterranean fruit fly *Ceratitis capitata* EcR ("CcEcR"), a locust *Locusta migratoria* EcR ("LmEcR"), an aphid *Myzus persicae* EcR ("MpEcR"), a fiddler crab *Uca pugilator* EcR ("UpEcR"), or an ixodid tick *Amblyomma americanum* EcR ("AmaEcR"). Even more preferably, the LBD is from spruce budworm (*Choristoneura fumiferana*) EcR ("CfEcR") or fruit fly *Drosophila melanogaster* EcR ("DmEcR").

[0159] Preferably, the LBD is from a truncated insect EcR. The insect EcR polypeptide truncation comprises a deletion of at least 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, 150, 155, 160, 165, 170, 175, 180, 195, 190, 195, 200, 205, 210, 215, 220, 225, 230, 235, 240, 245, 250, 255, 260, or 265 amino acids. More preferably, the insect EcR polypeptide truncation comprises a deletion of at least a partial polypeptide domain. Even more preferably, the insect EcR polypeptide truncation comprises a deletion of at least an entire polypeptide domain. In a specific embodiment, the insect EcR polypeptide truncation comprises a deletion of at least an A/B-domain deletion, a C-domain deletion, a D-domain deletion, an E-domain deletion, an F-domain deletion, an A/B/C-domains deletion, an A/B/I-2/C-domains deletion, an A/B/C/D-domains deletion, an A/B/C/D/F-domains deletion, an A/B/F-domains, and an A/B/C/F-domains deletion. A combination of several complete and/or partial domain deletions may also be performed.

[0160] In a preferred embodiment, the ecdysone receptor ligand binding domain is encoded by a polynucleotide comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, and SEQ ID NO: 10.

[0161] In another preferred embodiment, the ecdysone receptor ligand binding domain comprises an amino acid sequence selected from the group consisting of SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, and SEQ ID NO: 20.

[0162] Preferably, the RXR polypeptide is a mouse *Mus musculus* RXR ("MmRXR") or a human *Homo sapiens* RXR ("HsRXR"). The RXR polypeptide may be an RXR $\alpha$ , RXR $\beta$ , or RXR $\gamma$  isoform.

[0163] Preferably, the LBD is from a truncated RXR. The RXR polypeptide truncation comprises a deletion of at least 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, 150, 155, 160, 165, 170, 175, 180, 185, 190, 195, 200, 205, 210, 215, 220, 225, 230, 235, 240, 245, 250, 255, 260, or 265 amino acids. More preferably, the RXR polypeptide truncation comprises a deletion of at least a partial polypeptide domain. Even more preferably, the RXR polypeptide truncation comprises a deletion of at least an entire

polypeptide domain. In a specific embodiment, the RXR polypeptide truncation comprises a deletion of at least an A/B-domain deletion, a C-domain deletion, a D-domain deletion, an E-domain deletion, an F-domain deletion, an A/B/C-domains deletion, an AB1/2-C-domains deletion, an A/B/C/D-domains deletion, an A/B/C/D/P-domains deletion, an A/B/F-domains, and an A/B/C/F-domains deletion. A combination of several complete and/or partial domain deletions may also be performed.

[0164] In a preferred embodiment, the retinoid X receptor ligand binding domain is encoded by a polynucleotide comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, and SEQ ID NO: 30.

[0165] In another preferred embodiment, the retinoid X receptor ligand binding domain comprises an amino acid sequence selected from the group consisting of SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39, and SEQ ID NO: 40.

[0166] In a preferred embodiment, the gene expression cassette encodes a hybrid polypeptide comprising a DNA-binding domain encoded by a polynucleotide comprising a nucleic acid sequence selected from the group consisting of a GAL4 DBD (SEQ ID NO: 41) or a LexA DBD (SEQ ID NO: 43) and an ecdysone receptor ligand binding domain encoded by a polynucleotide comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, and SEQ ID NO: 10.

[0167] In another preferred embodiment, the gene expression cassette encodes a hybrid polypeptide comprising a DNA-binding domain comprising a polypeptide sequence selected from the group consisting of a GAL4 DBD (SEQ ID NO: 42) or a LexA DBD (SEQ ID NO: 44) and an ecdysone receptor ligand binding domain comprising an amino acid sequence selected from the group consisting of SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, and SEQ ID NO: 20.

[0168] In another preferred embodiment, the gene expression cassette encodes a hybrid polypeptide comprising a DNA-binding domain encoded by a polynucleotide comprising a nucleic acid sequence selected from the group consisting of a GAL4 DBD (SEQ ID NO: 41) or a LexA DBD (SEQ ID NO: 43) and a retinoid X receptor ligand binding domain encoded by a polynucleotide comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, and SEQ ID NO: 30.

[0169] In another preferred embodiment, the gene expression cassette encodes a hybrid polypeptide comprising a DNA-binding domain comprising a polypeptide sequence selected from the group consisting of a GAL4 DBD (SEQ ID NO: 42) or a LexA DBD (SEQ ID NO: 44) and a retinoid X receptor ligand binding domain comprising an amino acid sequence selected from the group consisting of SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39, and SEQ ID NO: 40.

[0170] In another preferred embodiment, the gene expression cassette encodes a hybrid polypeptide comprising a transactivation domain encoded by a polynucleotide comprising a nucleic acid sequence of SEQ ID NO: 45 and an ecdysone receptor ligand binding domain encoded by a polynucleotide comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, and SEQ ID NO: 10.

[0171] In another preferred embodiment, the gene expression cassette encodes a hybrid polypeptide comprising a transactivation domain comprising a polypeptide sequence of SEQ ID NO: 46 and an ecdysone receptor ligand binding domain comprising a polypeptide sequence selected from the group consisting of SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, and SEQ ID NO: 20.

[0172] In another preferred embodiment, the gene expression cassette encodes a hybrid polypeptide comprising a transactivation domain encoded by a polynucleotide comprising a nucleic acid sequence of SEQ ID NO: 45 and a retinoid X receptor ligand binding domain encoded by a polynucleotide comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, and SEQ ID NO: 30.

[0173] In another preferred embodiment, the gene expression cassette encodes a hybrid polypeptide comprising a transactivation domain comprising a polypeptide sequence of SEQ ID NO: 46 and a retinoid X receptor ligand binding domain comprising an amino acid sequence selected from the group consisting of SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO: 37, SEQ ID NO: 38, SEQ ID NO: 39, and SEQ ID NO: 40.

[0174] For purposes of this invention EcR and RXR also include synthetic and chimeric EcR and RXR and their homologs.

**POLYNUCLEOTIDES OF THE INVENTION**

**[0175]** The novel ecdysone receptor-based inducible gene expression system of the invention comprises a gene expression cassette comprising a polynucleotide that encodes a truncated EcR or RXR polypeptide comprising a truncation mutation and is useful in methods of modulating the expression of a gene within a host cell.

**[0176]** Thus, the present invention also relates to a polynucleotide that encodes an EcR or RXR polypeptide comprising a truncation mutation. Specifically, the present invention relates to an isolated polynucleotide encoding an EcR or RXR polypeptide comprising a truncation mutation that affects ligand binding activity or ligand sensitivity.

**[0177]** Preferably, the truncation mutation results in a polynucleotide that encodes a truncated EcR polypeptide or a truncated RXR polypeptide comprising a deletion of at least 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, 150, 155, 160, 165, 170, 175, 180, 185, 190, 195, 200, 205, 210, 215, 220, 225, 230, 235, 240, 245, 250, 255, 260, or 265 amino acids. More preferably, the EcR or RXR polypeptide truncation comprises a deletion of at least a partial polypeptide domain. Even more preferably, the EcR or RXR polypeptide truncation comprises a deletion of at least an entire polypeptide domain. In a specific embodiment, the EcR or RXR polypeptide truncation comprises a deletion of at least an A/B-domain deletion, a C-domain deletion, a D-domain deletion, an E-domain deletion, an F-domain deletion, an A/B/C-domains deletion, an A/B/C/D/F-domains deletion, an A/B/C/D/F-domains deletion, an A/B/F-domains, and an A/B/C/F-domains deletion. A combination of several complete and/or partial domain deletions may also be performed.

**[0178]** In a specific embodiment, the EcR polynucleotide according to the invention comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, and SEQ ID NO: 10. In a specific embodiment, the polynucleotide according to the invention encodes a ecdysone receptor polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO: 11 (CfEcR-CDEF), SEQ ID NO: 12 (CfEcR-1/2CDEF, which comprises the last 33 carboxy-terminal amino acids of C domain), SEQ ID NO: 13 (CfEcR-DEF), SEQ ID NO: 14 (CfEcR-E), SEQ ID NO: 15 (CfEcR-DE), SEQ ID NO: 16 (DmEcR-CDEF), SEQ ID NO: 17 (DmEcR-1/2CDEF), SEQ ID NO: 18 (DmEcR-DEF), SEQ ID NO: 19 (DmEcR-EF), and SEQ ID NO: 20 (DmEcR-E).

**[0179]** In another specific embodiment, the RXR polynucleotide according to the invention comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, and SEQ ID NO: 30. In another specific embodiment, the polynucleotide according to the invention encodes a truncated RXR polypeptide comprising an amino acid sequence consisting of SEQ ID NO: 31 (MmRXR-CDEF), SEQ ID NO: 32 (MmRXR-DEF), SEQ ID NO: 33 (MmRXR-EF), SEQ ID NO: 34 (MmRXR-truncatedEF), SEQ ID NO: 35 (MmRXR-E), SEQ ID NO: 36 (HsRXR-CDEF), SEQ ID NO: 37 (HsRXR-DEF), SEQ ID NO: 38 (HsRXR-EF), SEQ ID NO: 39 (HsRXR-truncated EF), and SEQ ID NO: 40 (HsRXR-E).

**[0180]** In particular, the present invention relates to an isolated polynucleotide encoding an EcR or RXR polypeptide comprising a truncation mutation, wherein the mutation reduces ligand binding activity or ligand sensitivity of the EcR or RXR polypeptide. In a specific embodiment, the present invention relates to an isolated polynucleotide encoding an EcR or RXR polypeptide comprising a truncation mutation that reduces steroid binding activity or steroid sensitivity of the EcR or RXR polypeptide. In a preferred embodiment, the present invention relates to an isolated polynucleotide encoding an EcR polypeptide comprising a truncation mutation that reduces steroid binding activity or steroid sensitivity of the EcR polypeptide, wherein the polynucleotide comprises a nucleic acid sequence of SEQ ID NO: 3 (CfEcR-DEF), SEQ ID NO: 4 (CfEcR-EF), SEQ ID NO: 8 (DmEcR-DEF), or SEQ ID NO: 9 (DmEcR-EF). In another specific embodiment, the present invention relates to an isolated polynucleotide encoding an EcR or RXR polypeptide comprising a truncation mutation that reduces non-steroid binding activity or non steroid sensitivity of the EcR or RXR polypeptide. In a preferred embodiment, the present invention relates to an isolated polynucleotide encoding an EcR or RXR polypeptide comprising a truncation mutation that reduces steroid binding activity or steroid sensitivity of the EcR or RXR polypeptide, wherein the polynucleotide comprises a nucleic acid sequence of SEQ ID NO: 4 (CfEcR-EF) or SEQ ID NO: 9 (DmEcR-EF).

**[0181]** The present invention also relates to an isolated polynucleotide encoding an EcR or RXR polypeptide comprising a truncation mutation, wherein the mutation enhances ligand binding activity or ligand sensitivity of the EcR or RXR polypeptide. In a specific embodiment, the present invention relates to an isolated polynucleotide encoding an EcR or RXR polypeptide comprising a truncation mutation that enhances steroid binding activity or steroid sensitivity of the EcR or RXR polypeptide. In another specific embodiment, the present invention relates to an isolated polynucleotide encoding an EcR or RXR polypeptide comprising a truncation mutation that enhances non-steroid binding activity or non-steroid sensitivity of the EcR or RXR polypeptide. In a preferred embodiment, the present invention relates to an isolated polynucleotide encoding an EcR polypeptide comprising a truncation mutation that enhances non-steroid binding activity or non-steroid sensitivity of the EcR polypeptide, wherein the polynucleotide comprises a nucleic acid sequence of SEQ ID NO: 3 (CfEcR-DEF) or SEQ ID NO: 8 (DmEcR-DEF).

**[0182]** The present invention also relates to an isolated polynucleotide encoding a retinoid X receptor polypeptide

comprising a truncation mutation that increases ligand sensitivity of a heterodimer comprising the mutated retinoid X receptor polypeptide and a dimerization partner. Preferably, the isolated polynucleotide encoding a retinoid X receptor polypeptide comprising a truncation mutation that increases ligand sensitivity of a heterodimer comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO: 23 (MmRXR-EF), SEQ ID NO: 24 (114mRXR-truncatedEF),  
 5 SEQ ID NO: 28 (HsRXR-EF), or SEQ ID NO: 29 (HsRXR-truncated EF). In a specific embodiment, the dimerization partner is an ecdysone receptor polypeptide. Preferably, the dimerization partner is a truncated EcR polypeptide. More preferably, the dimerization partner is an EcR polypeptide in which domains A/B/C have been deleted. Even more preferably, the dimerization partner is an EcR polypeptide comprising an amino acid sequence of SEQ ID NO: 13 (CfEcR-DEF) or SEQ ID NO: 18 (DmEcR-DEF).

10

## POLYPEPTIDES OF THE INVENTION

[0183] The novel ecdysone receptor-based inducible gene expression system of the invention comprises a polynucleotide that encodes a truncated EcR or RXR polypeptide and is useful in methods of modulating the expression of a gene within a host cell. Thus, the present invention also relates to an isolated truncated EcR or RXR polypeptide encoded by a polynucleotide or a gene expression cassette according to the invention. Specifically, the present invention relates to an isolated truncated EcR or RXR polypeptide comprising a truncation mutation that affects ligand binding activity or ligand sensitivity encoded by a polynucleotide according to the invention.

15

[0184] The present invention also relates to an isolated truncated EcR or RXR polypeptide comprising a truncation mutation. Specifically, the present invention relates to an isolated EcR or RXR polypeptide comprising a truncation mutation that affects ligand binding activity or ligand sensitivity.

20

[0185] Preferably, the truncation mutation results in a truncated EcR polypeptide or a truncated RXR polypeptide comprising a deletion of at least 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 105, 110, 115, 120, 125, 130, 135, 140, 145, 150, 155, 160, 165, 170, 175, 180, 185, 190, 195, 200, 205, 210, 215, 220, 225, 230, 235, 240, 245, 250, 255, 260, or 265 amino acids. More preferably, the EcR or RXR polypeptide truncation comprises a deletion of at least a partial polypeptide domain. Even more preferably, the EcR or RXR polypeptide truncation comprises a deletion of at least an entire polypeptide domain. In a specific embodiment, the EcR or RXR polypeptide truncation comprises a deletion of at least an A/B-domain deletion, a C-domain deletion, a D-domain deletion, an E-domain deletion, an F-domain deletion, an AB/C-domains deletion, an A/B/1/2-C-domains deletion, an AB/C/D-domains deletion, an AB/C/D/F-domains deletion, an A/B/F-domains, and an A/B/C/F-domains deletion. A combination of several complete and/or partial domain deletions may also be performed.

25

[0186] In a preferred embodiment, the isolated truncated ecdysone receptor polypeptide is encoded by a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO: 1 (CfEcR-CDEF), SEQ ID NO: 2 (CfEcR-1/2CDEF), SEQ ID NO: 3 (CfEcR-DEF), SEQ ID NO: 4 (CfEcR-EF), SEQ ID NO: 5 (CfEcR-DE), SEQ ID NO: 6 (DmEcR-CDEF), SEQ ID NO: 7 (DmEcR-1/2CDEF), SEQ ID NO: 8 (DmEcR-DEF), SEQ ID NO: 9 (DmEcR-EF), and SEQ ID NO: 10 (DmEcR-DE). In another preferred embodiment, the isolated truncated ecdysone receptor polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO: 11 (CfEcR-CDEF), SEQ ID NO: 12 (CfEcR-1/2CDEF), SEQ ID NO: 13 (CfEcR-DEF), SEQ ID NO: 14 (CfEcR-EF), SEQ ID NO: 15 (CfEcR-DE), SEQ ID NO: 16 (DmEcR-CDEF), SEQ ID NO: 17 (DmEcR-1/2CDEF), SEQ ID NO: 18 (DmEcR-DEF), SEQ ID NO: 19 (DmEcR-EF), and SEQ ID NO: 20 (DmEcR-DE).

30

[0187] In a preferred embodiment, the isolated truncated RXR polypeptide is encoded by a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO: 21 (MmRXR-CDEF), SEQ ID NO: 22 (MmRXR-DEF), SEQ ID NO: 23 (MmRXR-EF), SEQ ID NO: 24 (MmRXR-truncatedEF), SEQ ID NO: 25 (MmRXR-E), SEQ ID NO: 26 (HsRXR-CDEF), SEQ ID NO: 27 (HsRXR-DEF), SEQ ID NO: 28 (HsRXR-EF), SEQ ID NO: 29 (HsRXR-truncatedEF), and SEQ ID NO: 30 (HsRXR-E). In another preferred embodiment, the isolated truncated RXR polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO: 31 (MmRXR-CDEF), SEQ ID NO: 32 (MmRXR-DEF), SEQ ID NO: 33 (MmRXR-EF), SEQ ID NO: 34 (MmRXR-truncatedEF), SEQ ID NO: 35 (MmRXR-E), SEQ ID NO: 36 (HsRXR-CDEF), SEQ ID NO: 37 (HsRXR-DEF), SEQ ID NO: 38 (HsRXR-EF), SEQ ID NO: 39 (HsRXR-truncatedEF), and SEQ ID NO: 40 (HsRXR-E).

35

[0188] The present invention relates to an isolated EcR or RXR polypeptide comprising a truncation mutation that reduces ligand binding activity or ligand sensitivity of the EcR or RXR polypeptide, wherein the polypeptide is encoded by a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO: 1 (CfEcR-CDEF), SEQ ID NO: 2 (CfEcR-1/2CDEF), SEQ ID NO: 3 (CfEcR-DEF), SEQ ID NO: 4 (CfEcR-EF), SEQ ID NO: 5 (CfEcR-DE), SEQ ID NO: 6 (DmEcR-CDEF), SEQ ID NO: 7 (DmEcR-1/2CDEF), SEQ ID NO: 8 (DmEcR-DEF), SEQ ID NO: 9 (DmEcR-EF), SEQ ID NO: 10 (DmEcR-DE), SEQ ID NO: 21 (MmRXR-CDEF), SEQ ID NO: 22 (MmRXR-DEF), SEQ ID NO: 23 (MmRXR-EF), SEQ ID NO: 24 (MmRXR-truncatedEF), SEQ ID NO: 25 (MmRXR-E), SEQ ID NO: 26 (HsRXR-CDEF), SEQ ID NO: 27 (HsRXR-DEF), SEQ ID NO: 28 (HsRXR-EF), SEQ ID NO: 29 (HsRXR-truncatedEF), and SEQ ID NO: 30 (HsRXR-E).

40

45

50

55

**[0189]** Thus, the present invention relates to an isolated truncated EcR or RXR polypeptide comprising a truncation mutation that reduces ligand binding activity or ligand sensitivity of the EcR or RXR polypeptide, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO: 11 (CfEcR-CDEF), SEQ ID NO: 12 (CfEcR-1/2CDEF), SEQ ID NO: 13 (CfEcR-DEF), SEQ ID NO: 14 (CfEcR-EF), SEQ ID NO: 15 (CfEcR-DE), SEQ ID NO: 16 (DmEcR-CDEF), SEQ ID NO: 17 (DmEcR-1/2CDEF), SEQ ID NO: 18 (DmEcR-DEF), SEQ ID NO: 19 (DmEcR-EF), SEQ ID NO: 20 (DmEcR-DE), SEQ ID NO: 31 (MmRXR-CDEF), SEQ ID NO: 32 (MmRXR-DEF), SEQ ID NO: 33 (MmRXR-EF), SEQ ID NO: 34 (MmRXR-truncatedEF), SEQ ID NO: 35 (MmRXR-E), SEQ ID NO: 36 (HsRXR-CDEF), SEQ ID NO: 37 (HsRXR-DEF), SEQ ID NO: 38 (HsRXR-EF), SEQ ID NO: 39 (HsRXR-truncatedEF), and SEQ ID NO: 40 (HsRXR-E).

**[0190]** In a specific embodiment, the present invention relates to an isolated EcR or RXR polypeptide comprising a truncation mutation that reduces steroid binding activity or steroid sensitivity of the EcR or RXR polypeptide. In a preferred embodiment, the present invention relates to an isolated EcR polypeptide comprising a truncation mutation that reduces steroid binding activity or steroid sensitivity of the EcR polypeptide, wherein the EcR polypeptide is encoded by a polynucleotide comprising a nucleic acid sequence of SEQ ID NO: 3 (CfEcR-DEF), SEQ ID NO: 4 (CfEcR-EF), SEQ ID NO: 8 (DmEcR-DEF), or SEQ ID NO: 9 (DmEcR-EF). Accordingly, the present invention also relates to an isolated truncated EcR or RXR polypeptide comprising a truncation mutation that reduces steroid binding activity or steroid sensitivity of the EcR or RXR polypeptide. In a preferred embodiment, the present invention relates to an isolated EcR polypeptide comprising a truncation mutation that reduces steroid binding activity or steroid sensitivity of the EcR polypeptide, wherein the EcR polypeptide comprises an amino acid sequence of SEQ ID NO: 13 (CfEcR-DEF), SEQ ID NO: 14 (CfEcR-EF), SEQ ID NO: 18 (DmEcR-DEF), or SEQ ID NO: 19 (DmEcR-EF).

**[0191]** In another specific embodiment, the present invention relates to an isolated EcR or RXR polypeptide comprising a truncation mutation that reduces non-steroid binding activity or non-steroid sensitivity of the EcR or RXR polypeptide. In a preferred embodiment, the present invention relates to an isolated EcR polypeptide comprising a truncation mutation that reduces non-steroid binding activity or non-steroid sensitivity of the EcR polypeptide, wherein the EcR polypeptide is encoded by a polynucleotide comprising a nucleic acid sequence of SEQ ID NO: 4 (CfEcR-EF) or SEQ ID NO: 9 (DmEcR-EF). Accordingly, the present invention also relates to an isolated truncated EcR or RXR polypeptide comprising a truncation mutation that reduces non-steroid binding activity or steroid sensitivity of the EcR or RXR polypeptide. In a preferred embodiment, the present invention relates to an isolated EcR polypeptide comprising a truncation mutation that reduces non-steroid binding activity or non-steroid sensitivity of the EcR polypeptide, wherein the EcR polypeptide comprises an amino acid sequence of SEQ ID NO: 14 (CfEcR-EF) or SEQ ID NO: 19 (DmEcR-EF).

**[0192]** In particular, the present invention relates to an isolated EcR or RXR polypeptide comprising a truncation mutation that enhances ligand binding activity or ligand sensitivity of the EcR or RXR polypeptide, wherein the polypeptide is encoded by a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO: 1 (CfEcR-CDEF), SEQ ID NO: 2 (CfEcR-1/2CDEF), SEQ ID NO: 3 (CfEcR-DEF), SEQ ID NO: 4 (CfEcR-EF), SEQ ID NO: 5 (CfEcR-DE), SEQ ID NO: 6 (DmEcR-CDEF), SEQ ID NO: 7 (DmEcR-1/2CDEF), SEQ ID NO: 8 (DmEcR-DEF), SEQ ID NO: 9 (DmEcR-EF), SEQ ID NO: 10 (DmEcR-DE), SEQ ID NO: 21 (MmRXR-CDEF), SEQ ID NO: 22 (MmRXR-DEF), SEQ ID NO: 23 (MmRXR-EF), SEQ ID NO: 24 (MmRXR-truncatedEF), SEQ ID NO: 25 (MmRXR-E), SEQ ID NO: 26 (HsRXR-CDEF), SEQ ID NO: 27 (HsRXR-DEF), SEQ ID NO: 28 (HsRXR-EF), SEQ ID NO: 29 (HsRXR-truncatedEF), and SEQ ID NO: 30 (HsRXR-E).

**[0193]** The present invention relates to an isolated EcR or RXR polypeptide comprising a truncation mutation that enhances ligand binding activity or ligand sensitivity of the EcR or RXR polypeptide, wherein the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO: 11 (CfEcR-CDEF), SEQ ID NO: 12 (CfEcR-1/2CDEF), SEQ ID NO: 13 (CfEcR-DEF), SEQ ID NO: 14 (CfEcR-EF), SEQ ID NO: 15 (CfEcR-DE), SEQ ID NO: 16 (DmEcR-CDEF), SEQ ID NO: 17 (DmEcR-1/2CDEF), SEQ ID NO: 18 (DmEcR-DEF), SEQ ID NO: 19 (DmEcR-EF), SEQ ID NO: 20 (DmEcR-DE), SEQ ID NO: 31 (MmRXR-CDEF), SEQ ID NO: 32 (MmRXR-DEF), SEQ ID NO: 33 (MmRXR-EF), SEQ ID NO: 34 (MmRXR-truncatedEF), SEQ ID NO: 35 (MmRXR-E), SEQ ID NO: 36 (HsRXR-CDEF), SEQ ID NO: 37 (HsRXR-DEF), SEQ ID NO: 39 (HsRXR-EF), SEQ ID NO: 39 (HsRXR-truncatedEF), and SEQ ID NO: 40 (HsRXR-E).

**[0194]** The present invention relates to an isolated EcR or RXR polypeptide comprising a truncation mutation that enhances ligand binding activity or ligand sensitivity of the EcR or RXR polypeptide. In a specific embodiment, the present invention relates to an isolated EcR or RXR polypeptide comprising a truncation mutation that enhances steroid binding activity or steroid sensitivity of the EcR or RXR polypeptide. Accordingly, the present invention also relates to an isolated EcR or RXR polypeptide comprising a truncation mutation that enhances steroid binding activity or steroid sensitivity of the EcR or RXR polypeptide.

**[0195]** In another specific embodiment, the present invention relates to an isolated EcR or RXR polypeptide comprising a truncation mutation that enhances non-steroid binding activity or non-steroid sensitivity of the EcR or RXR polypeptide. In a preferred embodiment, the present invention relates to an isolated EcR polypeptide comprising a truncation mutation that enhances non-steroid binding activity or non-steroid sensitivity of the EcR polypeptide, wherein the EcR polypeptide

is encoded by a polynucleotide comprising a nucleic acid sequence of SEQ ID NO: 3 (CfEcR-DEF) or SEQ ID NO: 8 (DmEcR-DEF). Accordingly, the present invention also relates to an isolated EcR or RXR polypeptide comprising a truncation mutation that enhances non-steroid binding activity or steroid sensitivity of the EcR or RXR polypeptide. In a preferred embodiment, the present invention relates to an isolated EcR polypeptide comprising a truncation mutation that enhances non-steroid binding activity or non-steroid sensitivity of the EcR polypeptide, wherein the EcR polynucleotide comprises an amino acid sequence of SEQ ID NO: 13 (CfEcR-DEF) or SEQ ID NO: 18 (DmEcR-DEF).

**[0196]** The present invention also relates to an isolated retinoid X receptor polypeptide comprising a truncation mutation that increases ligand sensitivity of a heterodimer comprising the mutated retinoid X receptor polypeptide and a dimerization partner. Preferably, the isolated retinoid X receptor polypeptide comprising a truncation mutation that increases ligand sensitivity of a heterodimer is encoded by a polynucleotide comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO: 23 (MmRXR-EF), SEQ ID NO: 24 (MmRXR-truncatedEF), SEQ ID NO: 28 (HsRXR-EF), or SEQ ID NO: 29 (HsRXR-truncatedEF). More preferably, the isolated polynucleotide encoding a retinoid X receptor polypeptide comprising a truncation mutation that increases ligand sensitivity of a heterodimer comprises an amino acid sequence selected from the group consisting of SEQ ID NO: 33 (MmRXR-EF), SEQ ID NO: 34 (MmRXR-truncatedEF), SEQ ID NO: 38 (HsRXR-EF), or SEQ ID NO: 39 (HsRXR-truncatedEF).

**[0197]** In a specific embodiment, the dimerization partner is an ecdysone receptor polypeptide. Preferably, the dimerization partner is a truncated EcR polypeptide. More preferably, the dimerization partner is an EcR polypeptide in which domains A/B/C have been deleted. Even more preferably, the dimerization partner is an EcR polypeptide comprising an amino acid sequence of SEQ ID NO: 13 (CfEcR-DEF) or SEQ ID NO: 18 (DmEcR-DEF).

#### METHOD OF MODULATING GENE EXPRESSION OF THE INVENTION

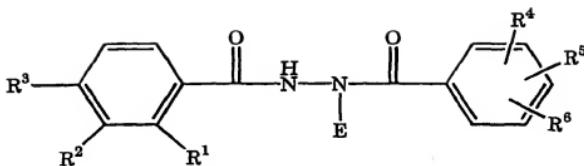
**[0198]** Applicants' invention also relates to methods of modulating gene expression in a host cell using a gene expression modulation system according to the invention. Specifically, Applicants' invention provides a method of modulating the expression of a gene in a host cell comprising the steps of: a) introducing into the host cell a gene expression modulation system according to the invention; and b) introducing into the host cell a ligand that independently combines with the ligand binding domains of the first polypeptide and the second polypeptide of the gene expression modulation system; wherein the gene to be expressed is a component of a gene expression cassette comprising: i) a response element comprising a domain to which the DNA binding domain of the first polypeptide binds; ii) a promoter that is activated by the transactivation domain of the second polypeptide; and iii) a gene whose expression is to be modulated, whereby a complex is formed comprising the ligand, the first polypeptide of the gene expression modulation system and the second polypeptide of the gene expression modulation system, and whereby the complex modulates expression of the gene in the host cell.

**[0199]** Genes of interest for expression in a host cell using Applicants' methods may be endogenous genes or heterologous genes. Nucleic acid or amino acid sequence information for a desired gene or protein can be located in one of many public access databases, for example, GENBANK, EMBL, Swiss-Prot, and PIR, or in many biology related journal publications. Thus, those skilled in the art have access to nucleic acid sequence information for virtually all known genes. Such information can then be used to construct the desired constructs for the insertion of the gene of interest within the gene expression cassettes used in Applicants' methods described herein.

**[0200]** Examples of genes of interest for expression in a host cell using Applicants' methods include, but are not limited to: antigens produced in plants as vaccines, enzymes like alpha-amylase, phytase, glucanases, and xylanase, genes for resistance against insects, nematodes, fungi, bacteria, viruses, and abiotic stresses, nutraceuticals, pharmaceuticals, vitamins, genes for modifying amino acid content, herbicide resistance, cold, drought, and heat tolerance, industrial products, oils, protein, carbohydrates, antioxidants, male sterile plants, flowers, fuels, other output traits, genes encoding therapeutically desirable polypeptides or products, such as genes that can provide, modulate, alleviate, correct and/or restore polypeptides important in treating a condition, a disease, a disorder, a dysfunction, a genetic defect, and the like.

**[0201]** Acceptable ligands are any that modulate expression of the gene when binding of the DNA binding domain of the two hybrid system to the response element in the presence of the ligand results in activation or suppression of expression of the genes. Preferred ligands include ponasterone, muristerone A, N,N'-diacylhydrazines such as those disclosed in U. S. Patents No. 6,013,836; 5,117,057; 5,530,028; and 5,378,726; dibenzoylalkyl cyanohydrazines such as those disclosed in European Application No. 461,809; N-alkyl-N,N'-diarylhydrazines such as those disclosed in U. S. Patent No. 5,225,443; N-acyl-N-alkylcarbonylhydrazines such as those disclosed in European Application No. 234,994; N-aryl-N-alkyl-N'-acylhydrazines such as those described in U. S. Patent No. 4,985,461; each of which is incorporated herein by reference and other similar materials including 3,5-di-tert-butyl-4-hydroxy-N-isobutyl-benzamide, 8-O-acetyl-harpagide, and the like.

**[0202]** Preferably, the ligand for use in Applicants' method of modulating expression of gene is a compound of the formula:



wherein:

E is a (C<sub>4</sub>-C<sub>6</sub>)alkyl containing a tertiary carbon or a cyano(C<sub>3</sub>-C<sub>5</sub>)alkyl containing a tertiary carbon;  
 15 R<sup>1</sup> is H, Me, Et, i-Pr, F, formyl, CF<sub>3</sub>, CUF<sub>2</sub>, CHCl<sub>2</sub>, CH<sub>2</sub>F, CH<sub>2</sub>Cl, CH<sub>2</sub>OH, CH<sub>2</sub>OMe, CH<sub>2</sub>CN, CN, C≡CH, 1-propynyl, 2-propynyl, vinyl, OH, OMe, OEt, cyclopropyl, CF<sub>2</sub>CF<sub>3</sub>, CH=CHCN, allyl, azido, SCN, or SCHF<sub>2</sub>;

R<sup>2</sup> is H, Me, Et, n-Pr, i-Pr, formyl, CF<sub>3</sub>, CHF<sub>2</sub>, CHCl<sub>2</sub>, CH<sub>2</sub>F, CH<sub>2</sub>Cl, CH<sub>2</sub>OH, CH<sub>2</sub>OMe, CH<sub>2</sub>CN, CN, C≡CH, 1-propynyl, 2-propynyl, vinyl, Ac, F, Cl, OH, OMe, OEt, O-n-Pr, OAc, NMe<sub>2</sub>, NEt<sub>2</sub>, SMe, SET, SOC<sub>2</sub>F, OCF<sub>2</sub>CF<sub>2</sub>H, COEt, cyclopropyl, CF<sub>2</sub>CF<sub>3</sub>, CH=CHCN, allyl, azido, OFC<sub>3</sub>, OCHF<sub>2</sub>, O-i-Pr, SCN, SCHF<sub>2</sub>, SMe, NH-CN, or joined with R<sup>3</sup> and the phenyl carbons to which R<sup>2</sup> and R<sup>3</sup> are attached to form an ethylenedioxy, a dihydrofuryl ring with the oxygen adjacent to a phenyl carbon, or a dihydropyrryl ring with the oxygen adjacent to a phenyl carbon;

R<sup>3</sup> is H, Et, or joined with R<sup>2</sup> and the phenyl carbons to which R<sup>2</sup> and R<sup>3</sup> are attached to form an ethylenedioxy, a dihydrofuryl ring with the oxygen adjacent to a phenyl carbon, or a dihydropyrryl ring with the oxygen adjacent to a phenyl carbon;

20 R<sup>4</sup>, R<sup>5</sup>, and R<sup>6</sup> are independently H, Me, Et, F, Cl, Br, formyl, CF<sub>3</sub>, CHF<sub>2</sub>, CHCl<sub>2</sub>, CH<sub>2</sub>F, CH<sub>2</sub>Cl, CH<sub>2</sub>OH, CN, C≡CH, 1-propynyl, 2-propynyl, vinyl, OMe, OEt, SMe, or SET.

30 [0203] Applicants' invention provides for modulation of gene expression in prokaryotic and eukaryotic host cells. Thus, the present invention also relates to a method for modulating gene expression in a host cell selected from the group consisting of a bacterial cell, a fungal cell, a yeast cell, a plant cell, an animal cell, and a mammalian cell. Preferably, the host cell is a yeast cell, a plant cell, a murine cell, or a human cell.

35 [0204] Expression in transgenic host cells may be useful for the expression of various polypeptides of interest including but not limited to therapeutic polypeptides, pathway intermediates; for the modulation of pathways already existing in the host for the synthesis of new products heretofore not possible using the host; cell based assays; and the like. Additionally the gene products may be useful for conferring higher growth yields of the host or for enabling alternative growth mode to be utilized.

#### 40 HOST CELLS AND NON-HUMAN ORGANISMS OF THE INVENTION

[0205] As described above, the gene expression modulation system of the present invention may be used to modulate gene expression in a host cell. Expression in transgenic host cells may be useful for the expression of various genes of interest. Thus, Applicants' invention also provides an isolated host cell comprising a gene expression system according to the invention. The present invention also provides an isolated host cell comprising a gene expression cassette according to the invention. Applicants' invention also provides an isolated host cell comprising a polynucleotide or polypeptide according to the invention. The isolated host cell may be either a prokaryotic or a eukaryotic host cell.

[0206] Preferably, the host cell is selected from the group consisting of a bacterial cell, a fungal cell, a yeast cell, a plant cell, an animal cell, and a mammalian cell. Examples of preferred host cells include, but are not limited to, fungal or yeast species such as *Aspergillus*, *Trichoderma*, *Saccharomyces*, *Pichia*, *Candida*, *Hansenula*, or bacterial species such as those in the genera *Synechocystis*, *Synechococcus*, *Salmonella*, *Bacillus*, *Acinetobacter*, *Rhodococcus*, *Streptomyces*, *Escherichia*, *Pseudomonas*, *Methyloimonas*, *Methyllobacter*, *Alcaligenes*, *Synechocystis*, *Anabaena*, *Thiobacillus*, *Methanobacterium* and *Klebsiella*, plant, animal, and mammalian host cells. More preferably, the host cell is a yeast cell, a plant cell, a mucine cell, or a human cell.

[0207] In a specific embodiment, the host cell is a yeast cell selected from the group consisting of a *Saccharomyces*, a *Pichia*, and a *Candida* host cell.

[0208] In another specific embodiment, the host cell is a plant cell selected from the group consisting of an apple, *Arabidopsis*, bajra, banana, barley, bean, beet, blackgram, chickpea, chili, cucumber, eggplant, favabean, maize, melon,

millet, mungbean, oat, okra, *Panicum*, papaya, peanut, pea, pepper, pigeonpea, pineapple, *Phaseolus*, potato, pumpkin, rice, sorghum, soybean, squash, sugarcane, sugarbeet, sunflower, sweet potato, tea, tomato, tobacco, watermelon, and wheat host cell.

[0209] In another specific embodiment, the host cell is a murine cell.

5 [0210] In another specific embodiment, the host cell is a human cell.

[0211] Host cell transformation is well known in the art and may be achieved by a variety of methods including but not limited to electroporation, viral infection, plasmid/vector transfection, non-viral vector mediated transfection, *Agrobacterium-mediated* transformation, particle bombardment, and the like. Expression of desired gene products involves culturing the transformed host cells under suitable conditions and inducing expression of the transformed gene. Culture conditions and gene expression protocols in prokaryotic and eukaryotic cells are well known in the art (see General Methods section of Examples). Cells may be harvested and the gene products isolated according to protocols specific for the gene product.

10 [0212] In addition, a host cell may be chosen which modulates the expression of the inserted polynucleotide, or modifies and processes the polypeptide product in the specific fashion desired. Different host cells have characteristic and specific mechanisms for the translational and post-translational processing and modification (e.g., glycosylation, cleavage [e.g., of signal sequence]) of proteins. Appropriate cell lines or host systems can be chosen to ensure the desired modification and processing of the foreign protein expressed. For example, expression in a bacterial system can be used to produce a non-glycosylated core protein product. However, a polypeptide expressed in bacteria may not be properly folded. Expression in yeast can produce a glycosylated product. Expression in eukaryotic cells can increase the likelihood of "native" glycosylation and folding of a heterologous protein. Moreover, expression in mammalian cells can provide a tool for reconstituting, or constituting, the polypeptide's activity. Furthermore, different vector/host expression systems may affect processing reactions, such as proteolytic cleavages, to a different extent.

15 [0213] Applicants' invention also relates to a non-human organism comprising an isolated host cell according to the invention. Preferably, the non-human organism is selected from the group consisting of a bacterium, a fungus, a yeast, a plant, an animal, and a mammal. More preferably, the non-human organism is a yeast, a plant, a mouse, a rat, a rabbit, a cat, a dog, a bovine, a goat, a pig, a horse, a sheep, a monkey, or a chimpanzee.

20 [0214] In a specific embodiment, the non-human organism is a yeast selected from the group consisting of *Saccharomyces*, *Pichia*, and *Candida*.

25 [0215] In another specific embodiment, the non-human organism is a plant selected from the group consisting of an apple, *Arabidopsis*, bairja, banana, barley, beans, beet, blackgram, chickpea, chili, cucumber, eggplant, favabean, maize, melon, millet, mungbean, oat, okra, *Panicum*, papaya, peanut, pea, pepper, pigeonpea, pineapple, *Phaseolus*, potato, pumpkin, rice, sorghum, soybean, squash, sugarcane, sugarbeet, sunflower, sweet potato, tea, tomato, tobacco, watermelon, and wheat.

30 [0216] In another specific embodiment, the non-human organism is a *Mus musculus* mouse.

#### **MEASURING GENE EXPRESSION/TRANSCRIPTION**

[0217] One useful measurement of Applicants' methods of the invention is that of the transcriptional state of the cell including the identities and abundances of RNA, preferably mRNA species. Such measurements are conveniently conducted by measuring cDNA abundances by any of several existing gene expression technologies.

40 [0218] Nucleic acid array technology is a useful technique for determining differential mRNA expression. Such technology includes, for example, oligonucleotide chips and DNA microarrays. These techniques rely on DNA fragments or oligonucleotides which correspond to different genes or cDNAs which are immobilized on a solid support and hybridized to probes prepared from total mRNA pools extracted from cells, tissues, or whole organisms and converted to cDNA. Oligonucleotide chips are arrays of oligonucleotides synthesized on a substrate using photolithographic techniques. Chips have been produced which can analyze for up to 1700 genes. DNA microarrays are arrays of DNA samples, typically PCR products, that are robotically printed onto a microscope slide. Each gene is analyzed by a full or partial-length target DNA sequence. Microarrays with up to 10,000 genes are now routinely prepared commercially. The primary difference between these two techniques is that oligonucleotide chips typically utilize 25-mer oligonucleotides which allow fractionation of short DNA molecules whereas the larger DNA targets of microarrays, approximately 1000 base pairs, may provide more sensitivity in fractionating complex DNA mixtures.

45 [0219] Another useful measurement of Applicants' methods of the invention is that of determining the translation state of the cell by measuring the abundances of the constituent protein species present in the cell using processes well known in the art.

50 [0220] Where identification of genes associated with various physiological functions is desired, an assay may be employed in which changes in such functions as cell growth, apoptosis, senescence, differentiation, adhesion, binding to a specific molecules, binding to another cell, cellular organization, organogenesis, intracellular transport, transport facilitation, energy conversion, metabolism, myogenesis, neurogenesis, and/or hematopoiesis is measured.

[0221] In addition, selectable marker or reporter gene expression may be used to measure gene expression modulation using Applicants' invention.

[0222] Other methods to detect the products of gene expression are well known in the art and include Southern blots (DNA detection), dot or slot blots (DNA, RNA), Northern blots (RNA), and RT-PCR (RNA) analyses. Although less preferred, labeled proteins can be used to detect a particular nucleic acid sequence to which it hybridizes.

[0223] In some cases it is necessary to amplify the amount of a nucleic acid sequence. This may be carried out using one or more of a number of suitable methods including, for example, polymerase chain reaction ("PCR"), ligase chain reaction ("LCR"), strand displacement amplification ("SDA"), transcription-based amplification, and the like. PCR is carried out in accordance with known techniques in which, for example, a nucleic acid sample is treated in the presence of a heat stable DNA polymerase, under hybridizing conditions, with one oligonucleotide primer for each strand of the specific sequence to be detected. An extension product of each primer that is synthesized is complementary to each of the two nucleic acid strands, with the primers sufficiently complementary to each strand of the specific sequence to hybridize therewith. The extension product synthesized from each primer can also serve as a template for further synthesis of extension products using the same primers. Following a sufficient number of rounds of synthesis of extension products, the sample may be analyzed as described above to assess whether the sequence or sequences to be detected are present.

[0224] The present invention may be better understood by reference to the following nonlimiting Examples, which are provided as exemplary of the invention.

## EXAMPLES

### GENERAL METHODS

[0225] Standard recombinant DNA and molecular cloning techniques used herein are well known in the art and are described by Sambrook, J., Fritsch, E. F. and Maniatis, T. *Molecular Cloning: A Laboratory Manual*; Cold Spring Harbor Laboratory Press; Cold Spring Harbor, (1989) (Maniatis) and by T. J. Silhavy, M. L. Bennan, and L. W. Enquist, *Experiments with Gene Fusions*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y. (1984) and by Ausubel, F. M. et al., *Current Protocols in Molecular Biology*, Greene Publishing Assoc. and Wiley-Interscience (1987).

[0226] Methods for plant tissue culture, transformation, plant molecular biology, and plant, general molecular biology may be found in *Plant Tissue Culture Concepts and Laboratory Exercises* edited by RN Trigiano and DJ Gray, 2<sup>nd</sup> edition, 2000, CRC press, New York; *Agrrobacterium Protocols* edited by KMA Garland and MR Davey, 1995, Humana Press, Totowa, New Jersey; *Methods in Plant Molecular Biology*, P. Maliga et al., 1995, Cold Spring Harbor Lab Press, New York; and *Molecular Cloning*, J. Sambrook et al., 1989, Cold Spring Harbor Lab Press, New York.

[0227] Materials and methods suitable for the maintenance and growth of bacterial cultures are well known in the art. Techniques suitable for use in the following examples may be found as set out in *Manual of Methods for General Bacteriology* (Philipp Gerhardt, R. G. E. Murray, Ralph N. Costilow, Eugene W. Nester, Willis A. Wood, Noel R. Krieg and G. Briggs Phillips, eds.), American Society for Microbiology, Washington, DC. (1994)) or by Thomas D. Brock in *Biotechnology: A Textbook of Industrial Microbiology*, Second Edition, Sinauer Associates, Inc., Sunderland, MA (1989). All reagents, restriction enzymes and materials used for the growth and maintenance of host cells were obtained from Aldrich Chemicals (Milwaukee, WI), DIFCO Laboratories (Detroit, MI), GIBCO/BRL (Gaithersburg, MD), or Sigma Chemical Company (St. Louis, MO) unless otherwise specified.

[0228] Manipulations of genetic sequences may be accomplished using the suite of programs available from the Genetics Computer Group Inc. (Wisconsin Package Version 9.0, Genetics Computer Group (GCG), Madison, WI). Where the GCG program "Pileup" is used the gap creation default value of 12, and the gap extension default value of 4 may be used. Where the GCG "Gap" or "Bestfit" programs is used the default gap creation penalty of 50 and the default gap extension penalty of 3 may be used. In any case where GCG program parameters are not prompted for, in these or any other GCG program, default values may be used.

[0229] The meaning of abbreviations is as follows: "h" means hour(s), "min" means minute(s), "sec" means second (s), "d" means day(s), " $\mu$ l" means microliter(s), "ml" means milliliter(s), "L" means liter(s), " $\mu$ M" means micromolar, "mM" means millimolar, " $\mu$ g" means microgram(s), "mg" means milligram(s), "A" means adenine or adenosine, "T" means thymine or thymidine, "G" means guanine or guanosine, "C" means cytidine or cytosine, "x g" means times gravity, "nt" means nucleotide(s), "aa" means amino acid(s), "bp" means base pair(s), "kb" means kilobase(s), "k" means kilo, " $\mu$ " means micro, and " $^{\circ}$ C" means degrees Celsius.

## EXAMPLE 1

[0230] Applicants' improved EcR-based inducible gene modulation system was developed for use in various applications including gene therapy, expression of proteins of interest in host cells, production of transgenic organisms, and

cell-based assays. This Example describes the construction and evaluation of several gene expression cassettes for use in the EcR-based inducible gene expression system of the invention.

**[0231]** In various cellular backgrounds, including mammalian cells, insect ecdysone receptor (EcR) heterodimerizes with retinoid X receptor (RXR) and, upon binding of ligand, transactivates genes under the control of ecdysone response elements. Applicants constructed several EcR-based gene expression cassettes based on the spruce budworm *Choristoneura fumiferana* EcR ("CfEcR"; full length polynucleotide and amino acid sequences are set forth in SEQ ID NO: 49 and SEQ ID NO: 50, respectively), *C. fumiferana* ultraspiracle ("CfUSP"; full length polynucleotide and amino acid sequences are set forth in SEQ ID NO: 51 and SEQ ID NO: 52, respectively), and mouse *Mus musculus* RXR $\alpha$  (MmRXR $\alpha$ ; full length polynucleotide and amino acid sequences are set forth in SEQ ID NO: 53 and SEQ ID NO: 54, respectively). The prepared receptor constructs comprise a ligand binding domain of EcR and of RXR or of USP; a DNA binding domain of GAL4 or of EcR; and an activation domain of VP16. The reporter constructs include a reporter gene, luciferase or LacZ, operably linked to a synthetic promoter construct that comprises either GAL4 or EcR/USP binding sites (response elements). Various combinations of these receptor and reporter constructs were cotransfected into CHO, NIH3T3, CV1 and 293 cells. Gene induction potential (magnitude of induction) and ligand specificity and sensitivity were examined using four different ligands: two steroidal ligands (ponasterone A and muristerone A) and two non-steroidal ligands (N-(2-ethyl-3-methoxybenzoyl)-N'-(3,5-dimethylbenzoyl)-N'-tert-butylhydrazine and N-(3,4-(1,2-ethylenedioxy)-2-methylbenzoyl)-N-(3,5-dimethylbenzoyl)-N'-tert-butylhydrazine) in a dose-dependent induction of reporter gene expression in the transfected cells. Reporter gene expression activities were assayed at 24hr or 48hr after ligand addition.

**Gene Expression Cassettes:** Ecdysone receptor-based, chemically inducible gene expression cassettes (switches) were constructed as follows, using standard cloning methods available in the art. The following is brief description of preparation and composition of each switch.

**1.1 - GAL4EcR/VP16RXR:** The D, E, and F domains from spruce budworm *Choristoneura fumiferana* EcR ("CfEcRDEF"; SEQ ID NO: 3) were fused to GAL4 DNA binding domain ("DNABD"; SEQ ID NO: 41) and placed under the control of an SV40e promoter (SEQ ID NO: 55). The DEF domains from mouse (*Mus musculus*) RXR ("MmRXRDEF"; SEQ ID NO: 22) were fused to the activation domain from VP16 ("VP16AD"; SEQ ID NO: 45) and placed under the control of an SV40e promoter (SEQ ID NO: 55). Five consensus GAL4 binding sites ("5XGAL4RE"; comprising 5, GAL4RE comprising SEQ ID NO: 47) were fused to a synthetic E1b minimal promoter (SEQ ID NO: 56) and placed upstream of the luciferase gene (SEQ ID NO: 57).

**1.2 - GAIAEcR/VP16USP:** This construct was prepared in the same way as in switch 1.1 above except MmRXR-DEF was replaced with the D, E and F domains from spruce budworm USP ("CfUSPDEF"; SEQ ID NO: 58). The constructs used in this example are similar to those disclosed in U. S. Patent No. 5,880,333 except that *Choristoneura fumiferana* USP rather than *Drosophila melanogaster* USP was utilized.

**1.3 - GAL4RXR/VP16CfEcR:** MmRXRDEF (SEQ ID NO: 22) was fused to a GAL4DNABD (SEQ ID NO: 41) and CfEcRCDEF (SEQ ID NO: 1) was fused to a VP16AD (SEQ ID NO: 45).

**1.4 - GAL4RXR/VP16DmEcR:** This construct was prepared in the same way as switch 1.3 except CfEcRCDEF was replaced with DmEcRCDEF (SEQ ID NO: 6).

**1.5 - GAL4USP/VP16CfEcR:** This construct was prepared in the same way as switch 1.3 except MmRXRDEF was replaced with CfUSPDEP (SEQ ID NO: 58).

**1.6 - GAIARXR(CfRPV16):** This construct was prepared so that both the GAL4 DNABD and the VP16AD were placed on the same molecule. GAL4DNABD (SEQ ID NO: 41) and VP16AD (SEQ ID NO: 45) were fused to CfEcRCDEF (SEQ ID NO: 3) at N- and C-termini respectively. The fusion was placed under the control of an SV40e promoter (SEQ ID NO: 55).

**1.7 - VP16CfEcR:** This construct was prepared such that CfEcRCDEF (SEQ ID NO: 1) was fused to VP16AD (SEQ ID NO: 45) and placed under the control of an SV40e promoter (SEQ ID NO: 55). Six ecdysone response elements ("EcRE"; SEQ ID NO: 59) from the hsp27 gene were placed upstream of the promoter and a luciferase gene (SEQ ID NO: 57). This switch most probably uses endogenous RXR.

**1.8 - DmVgRXR:** This system was purchased from Invitrogen Corp., Carlsbad, California. It comprises a *Drosophila melanogaster* EcR ("DmEcR") with a modified DNABD fused to VP16AD and placed under the control of a CMV promoter (SEQ ID NO: 60). Full length MmRXR (SEQ ID NO: 53) was placed under the control of the RSV promoter (SEQ ID NO: 61). The reporter, pIND(SP1)LacZ, contains five copies of a modified ecdysone response element ("EcRE", E/GRE), three copies of an SP1 enhancer, and a minimal heat shock promoter, all of which were placed upstream to the LacZ reporter gene.

**1.9 - CNgRXR:** This example was prepared in the same way as switch 1.8 except DmEcR was replaced with a truncated CfEcR comprising a partial A/B domain and the complete CDEF domains [SEQ ID NO: 62 (polynucleotide) and SEQ ID NO: 63 (polypeptide)].

**1.10 - CtVgRXRdel:** This example was prepared in the same way as switch 1.9 except MmRXR (SEQ ID NO:

53) was deleted.

Cell lines: Four cell lines: CHO, Chinese hamster *Cricetulus griseus* ovarian cell line; NIH3T3 (3T3) mouse *Mus musculus* cell line; 293 human *Homo sapiens* kidney cell line, and CV1 African green monkey kidney cell line were used in these experiments. Cells were maintained in their respective media and were subcultured when they reached 60% confluence. Standard methods for culture and maintenance of the cells were followed.

Transfections: Several commercially available lipofactors as well as electroporation methods were evaluated and the best conditions for transfection of each cell line were developed. CHO, NIH3T3, 293 and CV1 cells were grown to 60% confluence. DNAs corresponding to the various switch constructs outlined in Examples 1.1 through 1.10 were transfected into CHO cells, NIH3T3 cells, 293 cells, or CV1 cells as follows.

**CHO cells:** Cells were harvested when they reach 60-80% confluence and plated in 6- or 12- or 24- well plates at 250,000, 100,000, or 50,000 cells in 2.5, 1.0, or 0.5 ml of growth medium containing 10% Fetal bovine serum respectively. The next day, the cells were rinsed with growth medium and transfected for four hours. LipofectAMINE™ 2000 (Life Technologies Inc.) was found to be the best transfection reagent for these cells. For 12- well plates, 4  $\mu$ l of LipofectAMINE™ 2000 was mixed with 100  $\mu$ l of growth medium 1.0  $\mu$ g of reporter construct and 0.25  $\mu$ g of receptor construct(s) were added to the transfection mix. A second reporter construct was added [pTKRL (Promega, 0.1  $\mu$ g/transfection mix] and comprised a *Renilla* luciferase gene (SEQ ID NO: 64) operably linked and placed under the control of a thymidine kinase (TK) constitutive promoter and was used for normalization. The contents of the transfection mix were mixed in a vortex mixer and let stand at room temperature for 30 min. At the end of incubation, the transfection mix was added to the cells maintained in 400  $\mu$ l growth medium. The cells were maintained at 37 °C and 5% CO<sub>2</sub> for four hours. At the end of incubation, 500  $\mu$ l of growth medium containing 20% FBS and either DMSO (control) or a DMSO solution of appropriate ligands were added and the cells were maintained at 37 °C and 5% CO<sub>2</sub> for 24-48 hr. The cells were harvested and reporter activity was assayed. The same procedure was followed for 6 and 24 well plates as well except all the reagents were doubled for 6 well plates and reduced to half for 24-well plates. —

**NIH3T3 Cells:** Superfect™ (Qiagen Inc.) was found to be the best transfection reagent for 3T3 cells. The same procedures described for CHO cells were followed for 3T3 cells as well with two modifications. The cells were plated when they reached 50% confluence, 125,000 or 50,000 or 25,000 cells were plated per well of 6- or 12- or 24-well plates respectively. The GA14EcR/VP16RXR and reporter vector DNAs were transfected into NIH3T3 cells, the transfected cells were grown in medium containing PonA, MurA, N-(2-ethyl-3-methoxybenzoyl)-N'-(3,5-dimethylbenzoyl)-N'-t-butylhydrazine, or N-(3,4-(1,2-ethylenedioxy)-2-methylbenzoyl)-N-(3,5-dimethylbenzoyl)-N-tert-butylhydrazine for 48 hr. The ligand treatments were performed as described in the CHO cell section above.

**293 Cells:** LipofectAMINE™ 2000 (Life Technologies) was found to be the best lipofactor for 293 cells. The same procedures described for CHO were followed for 293 cells except that the cells were plated in biocoated plates to avoid clumping. The ligand treatments were performed as described in the CHO cell section above.

**CV1 Cells:** LipofectAMINE™ plus (Life Technologies) was found to be the best lipofactor for CV1 cells. The same procedures described for NIH3T3 cells were followed for CV1 cells

**Ligands:** Ponasterone A and Muristerone A were purchased from Sigma Chemical Company. The two non-steroids N-(2-ethyl-3-methoxybenzoyl)-N'-(3,5-dimethylbenzoyl)-N'-t-butylhydrazine, or N-(3,4-(1,2-ethylenedioxy)-2-methylbenzoyl)-N'-(3,5-dimethylbenzoyl)-N-tert-butylhydrazine are synthetic stable ecdysteroids synthesized at Rohm and Haas Company. All ligands were dissolved in DMSO and the final concentration of DMSO was maintained at 0.1 % in both controls and treatments.

**Reporter Assays:** Cells were harvested 24-48 hr after adding ligands. 125, 250, or 500  $\mu$ l of passive lysis buffer (part of Dual-luciferase™ reporter assay system from Promega Corporation) were added to each well of 24- or 12- or 24-well plates respectively. The plates were placed on a rotary shaker for 15 min. Twenty  $\mu$ l of lysate was assayed. Luciferase activity was measured using Dual-luciferase™ reporter assay system from Promega Corporation following the manufacturer's instructions.  $\beta$ -Galactosidase was measured using Galacto-Star™ assay kit from TROPIX following the manufacturer's instructions. All luciferase and  $\beta$ -galactosidase activities were normalized using *Renilla* luciferase as a standard. Fold activities were calculated by dividing normalized relative light units ("RLU") in ligand treated cells with normalized RLU in DMSO treated cells (untreated control).

[0232] The results of these experiments are provided in the following tables.

Table 1

Transactivation of reporter genes through various switches in CHO cells		
	Composition of Switch	Mean Fold Activation with 50µM N-(2-ethyl-3-methoxybenzoyl)-N'-(3,5-dimethylbenzoyl)-N-t-butylhydrazine
5	1.1 GAL4EcR + VP16RXR pGAL4RELuc	267
10	1.2 GAL4EcR + VP16USP pGAL4RELuc	2
15	1.3 GAL4RXR + VP16CfEcR pGAL4RELuc	85
20	1.4 GAL4RXR+VP16DmEcR pGAL4RELuc	312
25	1.5 GAL4USP + VP16CfEcR pGAL4RELuc	2
30	1.6 GAL4CfEcRVP16 pGAL4RELuc	9
	1.7 VP16CfEcR pEcRELuc	36
	1.8 DmVgRXR+MmRXR pIND(SP1)LacZ	14
	1.9 CfVgRXR + MmRXR pIND(SP1)LacZ	27
	1.10 CfVgRXR pIND(SP1)LacZ	29

Table 2

Transactivation of reporter genes through various switches in 3T3 cells		
	Composition of Switch	Mean Fold Activation Through N-(2-ethyl-3-methoxybenzoyl)-N'-(3,5-dimethylbenzoyl)-N-t-butylhydrazine
35	1.1 GAL4EcR + VP16RXR pGAL4RELuc	1118
40	1.2 GAL4EcR + VP16USP pGAL4RELuc	2
45	1.3 GAL4RXR + VP16CfEcR pGAL4RELuc	47
50	1.4 GAL4RXR + VP16DmEcR pGAL4RELuc	269
55	1.5 GAL4USP + VP16CfEcR pGAL4RELuc	3
	1.6 GAL4CfEcRVP16 pGAL4RELuc	7
	1.7 VP16CfEcR pEcRELuc	1
	1.8 DmVgRXR+MmRXR pIND(SP1)LacZ	21
	1.9 CfVgRXR + MmR7Gt pIND(SP1)LacZ	19
	1.10 CfVgRXR pIND(SP1)LacZ	2

(continued)

<b>Transactivation of reporter genes through various switches in 3T3 cells</b>	
Composition of Switch	Mean Fold Activation Through N-(2-ethyl-3-methoxybenzoyl)-N'-(3,5-dimethylbenzoyl)-N'-t-butylhydrazine
pIND(SP1)LacZ	

Table 3

<b>Transactivation of reporter genes through various switches in 293 cells</b>	
Composition of Switch	Mean Fold Activation Through N-(2-ethyl-3-methoxybenzoyl)-N'-(3,5-dimethylbenzoyl)-N'-t-butylhydrazine
1.1 GAL4EcR+VP16RXR pGAL4RELuc	125
1.2 GAL4EcR + VP16USP pGAL4RELuc	2
1.3 GAL4RXR+VP16CfEcR pGAL4RELuc	17
1.4 GAL4RXR + VP16DmEcR pGAL4RELuc	3
1.5 GAL4USP + VP16CfEcR pGAL4RELuc	2
1.6 GAL4CfEcRVP16 pGAL4RELuc	3
1.7 VP16CfEcR pEcRELuc	2
1.8 DmVgRXR+ MmRXR pIND(SP1)LacZ	21
1.9 CfVgRXR + MuaxRXR pIND(SP1)LacZ	12
1.10 CfVgRXR pIND(SP1)LacZ	3

Table 4

<b>Transactivation of reporter genes through various switches in CV1 cells</b>	
Composition of Switch	Mean Fold Activation Through N-(2-ethyl-3-methoxybenzoyl)-N'-(3,5-dimethylbenzoyl)-N'-t-butylhydrazine
1.1 GAL4EcR + VP16RXR pGAL4RELuc	279
1.2 GAL4EcR+VP16USP pGAL4RELuc	2
1.3 GAL4RXR + VP16CfEcR pGAL4RELuc	25
1.4 GAL4RXR + VP16DmEcR pGAL4RELuc	80
1.5 GAL4USP + VP16CfEcR pGAL4RELuc	3
1.6 GAL4CfEcRVP16 pGAL4RELuc	6

(continued)

Transactivation of reporter genes through various switches in CV1 cells		
	Composition of Switch	Mean Fold Activation Through N-(2-ethyl-3-methoxybenzoyl)-N'-(3,5-dimethylbenzoyl)-N'-t-butylhydrazine
5	1.7 VP16CfEcR	1
	pEcRELuc	
10	1.8 DmVgRXR+MmRXR	12
	pIND(SP1)LacZ	
	1.9 CfVgRXR + MmRXR	7
	pIND(SP1)LacZ	
15	1.10 Cf VgRXR	1
	pIND(SP1)LacZ	

Table 5

Transactivation of reporter gene GAL4CfEcRDEF/VP16MmRXRDEF (switch 1.1) through steroids and non-steroids in 3T3 cells.		
	Ligand	Mean Fold Induction at 1.0 µM Concentration
20	1. Ponasterone A	1.0
	2. Muristerone A	1.0
25	3. N-(2-ethyl-3-methoxybenzoyl)-N'-(3,5-dimethylbenzoyl)-N'-tert-butylhydrazine	116
	4. N'-(3,4-(1,2-ethylenedioxy)-2-methylbenzoyl)-N'-(3,5-diunethylbenzoyl)-N'-tert butylhydrazine	601

Table 6

Transactivation of reporter gene GAL4MmRXRDEF/VP16CfEcRCDEF (switch 1.3) through steroids and non-steroids in 3T3 cells.		
	Ligand	Mean Fold Induction at 1.0 µM Concentration
35	1. Ponasterone A	1.0
	2. Muristerone A	1.0
40	3. N-(2-ethyl-3-methoxybenzoyl)-N'-(3,5-dimethylbenzoyl)-N'-tert-butylhydrazine	71
	4. N'-(3,4-(1,2-ethylenedioxy)-2-methylbenzoyl)-N'-(3,5-dimethylbenzoyl)-N'-tert-butylhydrazine	54

[0233] Applicants' results demonstrate that the non-steroidal ecdysone agonists, N-(2-ethyl-3-methoxybenzoyl)-N'-(3,5-dimethylbenzoyl)-N'-tert-butylhydrazine and N'-(3,4-(1,2-ethylenedioxy)-2-methylbenzoyl)-N'-(3,5-dimethylbenzoyl)-N'-tert-butylhydrazine, were more potent activators of CfEcR as compared to *Drosophila melanogaster* EcR (DmEcR). (see Tables 1-4). Also, in the mammalian cell lines tested, MmRXR performed better than CfUSP as a heterodimeric partner for CfEcR. (see Tables 1-4). Additionally, Applicants' inducible gene expression modulation system performed better when exogenous MmRXR was used than when the system relied only on endogenous RXR levels (see Tables 1-4).

[0234] Applicants' results also show that in a CfEcR-based inducible gene expression system, the non-steroidal ecdysone agonists induced reporter gene expression at a lower concentration (i.e., increased ligand sensitivity) as compared to the steroid ligands, ponasterone A and muristerone A (see Tables 5 and 6).

[0235] Out of 10 EcR based gene switches tested, the GAL4EcR/VP16RXR switch (Switch 1.1) performed better than any other switch in all four cell lines examined and was more sensitive to non-steroids than steroids. The results also demonstrate that placing the activation domain (AD) and DNA binding domain (DNABD) on each of the two partners reduced background when compared to placing both AD and DNABD together on one of the two partners. Therefore,

a switch format where the AD and DNABD are separated between two partners, works well for EcR-based gene switch applications.

[0236] In addition, the MmRXR/EcR-based switches performed better than CfUSP/EcR-based switches, which have a higher background activity than the MmRXR/EcR switches in the absence of ligand.

[0237] Finally, the GAL4EcR/VP16RXR switch (Switch 1.1) was more sensitive to non-steroid ligands than to the steroid ligands (see Tables 5 and 6). In particular, steroid ligands initiated transactivation at concentrations of 50  $\mu$ M, whereas the non-steroid ligands initiated transactivation at less than 1  $\mu$ M (submicromolar) concentration.

## EXAMPLE 2

[0238] This Example describes Applicants' further analysis of truncated EcR and RXR polypeptides in the improved EcR-based inducible gene expression system of the invention. To identify the best combination and length of two receptors that give a switch with a) maximum induction in the presence of ligand; b) minimum background in the absence of ligand; c) highly sensitive to ligand concentration; and d) minimum cross-talk among ligands and receptors, Applicants made and analyzed several truncation mutations of the CfEcR and MmRXR receptor polypeptides in NIH3T3 cells.

[0239] Briefly, polynucleotides encoding EcR or RXR receptors were truncated at the junctions of A/B, C, D, E and F domains and fused either a GAL4 DNA binding domain encoding polynucleotide (SEQ ID NO: 41) for CfEcR, or a VP16 activation domain encoding polynucleotide (SEQ ID NO: 45) for MmRXR as described in Example 1. The resulting receptor truncation/fusion polypeptides were assayed in NIH3T3 cells. Plasmid pFRLUC (Stratagene) encoding a luciferase polypeptide was used as a reporter gene construct and pTKRL (Promega) encoding a *Renilla* luciferase polypeptide under the control of the constitutive TK promoter was used to normalize the transfections as described above. The analysis was performed in triplicates and mean luciferase counts were determined as described above.

### Gene Expression Cassettes Encoding Truncated Ecdysone Receptor Polypeptides

[0240] Gene expression cassettes comprising polynucleotides encoding either full length or truncated CfEcR polypeptides fused to a GAL4 DNA binding domain (SEQ ID NO: 41): GAL4CfEcRABCDEF (full length CfEcRA/BCDEF; SEQ ID NO: 49), GAL4CfEcRCDEF (CfEcRCDEF; SEQ ID NO: 1), GAL4CfEcR1/2CDEF (CfEcR1/2CDEF; SEQ ID NO: 2), GAL4CfEcRDEF (CfEcRDEF; SEQ ID NO: 3), GAL4CfEcREF (CfEcREF; SEQ ID NO: 4), and GAL4CfEcRDE (CfEcRDE; SEQ ID NO: 5) were transfected into NIH3T3 cells along with VP16MmRXRDEF (constructed as in Example 1.1; Figure 11) or VP16MmRXREF [constructed as in Example 1.1 except that MmRXRDEF was replaced with MmRXREF (SEQ ID NO: 23); Figure 12], and pFRLuc and pTKRL plasmid DNAs. The transfected cells were grown in the presence 0, 1.5 or 25  $\mu$ M of N-(2-ethyl-3-methoxybenzoyl)-N'-(3,5-dimethylbenzoyl)-N'-tert-butylhydrazine or PonA for 48 hr. The cells were harvested, lysed and luciferase reporter activity was measured in the cell lysates. Total fly luciferase relative light units are presented. The number on the top of each bar is the maximum fold induction for that treatment.

[0241] Applicants' results show that the EF domain of MmRXR is sufficient and performs better than DEF domains of this receptor (see Figures 11 and 12). Applicants have also shown that, in general, EcR/RXR receptor combinations are insensitive to PonA (see Figures 11 and 12). As shown in the Figures 11 and 12, the GAL4CfEcRCDEF hybrid polypeptide (SEQ ID NO: 7) performed better than any other CfEcR hybrid polypeptide.

### Gene Expression Cassettes Encoding Truncated Retinoid X Receptor Polypeptides

[0242] Gene expression cassettes comprising polynucleotides encoding either full length or truncated MmRXR polypeptides fused to a VP16 transactivation domain (SEQ ID NO: 45): VP16MmRXRA/BCDEF (full length MmRXRA/BCDEF; SEQ ID NO: 53), VP16MmRXRCDEF (MmRXRCDEF; SEQ ID NO: 21), VP16MmRXRDEF (MmRXR3CRDEF; SEQ ID NO: 22), VP16MmRXREF (MmRXREF; SEQ ID NO: 23), VP16MmRXRBam-EF ("MmRXRBam-EF" or "MmRXR-truncatedEF"; SEQ ID NO: 24), and VP16MmRXRAF2de1 ("MmRXRAF2de1" or "MmRXR-E"; SEQ ID NO: 25) constructs were transfected into NIH3T3 cells along with GAL4CfEcRCDEF (constructed as in Example 1.1; Figure 13) or GAL4CfEcRDEF [constructed as in Example 1.1 except CfEcRCDEF was replaced with CfEcRDEF (SEQ ID NO: 3; Figure 14)], pFRLuc and pTKRL plasmid DNAs as described above. The transfected cells were grown in the presence 0, 1, 5 and 25  $\mu$ M of N-(2-ethyl-3-methoxybenzoyl)-N'-(3,5-dimethylbenzoyl)-N'-tert-butylhydrazine or PonA for 48 hr. The cells were harvested and lysed and reporter activity was measured in the cell lysate. Total fly luciferase relative light units are presented. The number on top of each bar is the maximum fold induction in that treatment.

[0243] Of all the truncations of MmRXR tested, Applicants' results show that the MmRXREF receptor was the best partner for CfEcR (Figures 13 and 14). CfEcRCDEF showed better induction than CfEcRDEF using MmRXREF. Deleting AF2 (abbreviated "EF-AF2de1") or helices 1-3 of the E domain (abbreviated "EF-Bamdel") resulted in an RXR receptor that reduced gene induction and ligand sensitivity when partnered with either CfEcRCDEF (Figure 13) or CfEcRDEF (Figure 14) in NIH3T3 cells. In general, the CfEcR/RXR-based switch was much more sensitive to the non-steroid N-(2-

ethyl-3-methoxybenzoyl)-N'-(3,5-dimethylbenzoyl)-N'-tert-butylhydrazine than to the steroid PonA.

### EXAMPLE 3

[0244] This Example describes Applicants' further analysis of gene expression cassettes encoding truncated EcR or RXR receptor polypeptides that affect either ligand binding activity or ligand sensitivity, or both. Briefly, six different combinations of chimeric receptor pairs, constructed as described in Examples 1 and 2, were further analyzed in a single experiment in NIH3T3 cells. These six receptor pair combinations and their corresponding sample numbers are depicted in Table 7.

10

**Table 7**

CfEcR + MmRXR Truncation Receptor Combinations in NIH3T3 Cells		
Figure 15 X-Axis Sample No.	EcR Polypeptide Construct	RXR Polypeptide Construct
Samples 1 and 2	GAL4CfEcRCDEF	VP16RXRABCDEF (Full length)
Samples 3 and 4	GAL4CfEcRCDEF	VP16RXRDEF
Samples 5 and 6	GAL4CfEcRCDEP	VP16RXREF
Samples 7 and 8	GAL4CfEcRDEF	VP16RXRA/BCDEF (Full length)
Samples 9 and 10	GAL4CfEcRDEF	VP16RXRDEF
Samples 11 and 12	GAL4CfEcRDEF	VP16RXREF

[0245] The above receptor construct pairs, along with the reporter plasmid pFRLuc were transfected into NIH3T3 cells as described above. The six CfEcR truncation receptor combinations were duplicated into two groups and treated with either steroid (odd numbers on x-axis of Figure 15) or non-steroid (even numbers on x-axis of Figure 15). In particular, the cells were grown in media containing 0, 1, 5 or 25 uM PonA (steroid) or N-(2-ethyl-3-methoxybenzoyl)-N'-(3,5-dimethylbenzoyl)-N'-tert-butylhydrazine (non-steroid) ligand. The reporter gene activity was measured and total RLU are shown. The number on top of each bar is the maximum fold induction for that treatment and is the mean of three replicates.

[0246] As shown in Figure 15, the CfEcRCDEF/MmRXREF receptor combinations were the best switch pairs both in terms of total RLU and fold induction (compare columns 1-6 to columns 7-12). This confirms Applicants' earlier findings as described in Example 2 (Figures 11-14). The same gene expression cassettes encoding the truncated EcR and RXR polypeptides were also assayed in a human lung carcinoma cell line A549 (ATCC) and similar results were observed (data not shown).

### SEQUENCE LISTING

#### [0247]

40

<110> Rohm and Haas Company  
 Palli, Subba Reddy  
 Kapitskaya, Marianna Zinovjevna  
 Cress, Dean Ervin  
 <120> Novel Ecdysone Receptor-Based Inducible Gene Expression System  
 <130> RH0020  
 <150> 60/191,355  
 <151> 2000-03-22  
 <150> 60/269,799  
 <151> 2001-02-20  
 <160> 64  
 <170> PatentIn version 3.0 ,  
 <210> 1  
 <211> 1288  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <221> misc feature

&lt;223&gt; Novel Sequence

&lt;400&gt; 1

5	aaggccccgtca gcaagaggaa ctgtgtctgg tatgcgggga cagagcctcc ggataccact acaaatgcgt cacgtgtaa gggtgtaaag gtttcttcag acggagtgtt	60 120
10	acccaaaaatg cggtttataat ttgttaattc ggtcacgctt gcgaaaatggaa catgtacatcg cgacggaaaat gccaggagtg ccgcctgtaa aagtgtcttag ctgttaggcat gaggccttag	180 240
15	tgcgttagtac ccgagactca gtgcgtccat aagcggaaag agaagaaaagc acagaaggag aaggacaaac tgcgtgtcag cacgtacgac gtggacgacc acatgccg cattatcgag	300 360
20	tgtgaaacctc caccctctga agcagcaagg attcacgaag tggtcccaag gtttcttc gacaagctgt tggagacaaa cccgcagaaa aacatcccc agttgacagc caaccagcag	420 480
25	ttcccttatcg ccaggtcat ctggtaccag gacgggtacg acgacgcctc tgatgaagat ttgaagagga ttacgcacac gtggcagcaa gcggcagatg aaaacgaaaga gtcgtacact	540 600
30	cccttccgcg agatcacaga gatgactatc ctcacggtcc aacttacgtt ggagttcgcg aaggattgc cagggttcgc caagatctcg cagcctgtac aaattacgtc gcttaaggt	660 720
35	tgctcaagtg aggtaatgtat gtcggcgtc ggcgcacat acgtgcggc ctcagacagt gttctgttc cgaacaaacca aegtcaact cggcacaact accgcacggc tggcatggcc tacgtcatacg aggatctact gcaactctgc cgggtcatgt actctatggc gttggacaaac	780 840
40	atccattaaacg cgctgtctac ggctgtctgc atcttttctg accggccagg gttggagcag ccgcaactgg tggaaagaaat ccagcggtac tacgtataa cgtccgcat ctatatctcg	900 960
	aaccagctga gcgggtcggc gcggtcgtcc gtcataatacg gcaagatcct ctcaatcctc tctggactac gcacgcgtcg catgaaaac tccaaatgtt gcatctccat caaggtcaag	1080 1140
	aacagaaaacg tgccgcctttt cctcgaggag atctggatg tggcggacat gtcgcacacc caaccgcgcg cttatctcgaa gtcacccacg aatctctagc cctgcgcgc acgcatacgcc	1200 1260
	gatgcgcgtt ccggccgcgc tgctctga	1288

&lt;210&gt; 2

&lt;211&gt; 1110

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Novel Sequence

&lt;400&gt; 2

	gcggttata tttgtaaatt cggtcacgct tgcgaaatgg acatgtacat gcgacggaaa	60
5	tgccaggagt gccgcctgaa gaagtgccta gctgttaggc tgaggcctga gtgcgttagta	120
	cccgagactc agtgcgcctat gaagcgaaaa gagaagaaag cacagaagga gaaggacaaa	180
	ctgcctgtca gcacgacgac ggtggacgac cacatgccgc ccattatgcg gtgtgaacct	240
	ccacctctg aagcagcaag gattcacgaa gtggtcccaa ggtttctctc cgacaagctg	300
10	ttggagacaa accggcagaa aaacatcccc cagttgacag ccaaccagca gttecttata	360
	gccaggctca tctggatcca ggacgggtac gagcagcctt ctgatgaga ttgaagagg	420
	attacgcaga cgtgcgcaga akgcggacgat gaaaacgaaag agtctgacac tcccttccgc	480
15	cagatcaca agatgactat ctcacggc tc caacttatcg tggaggctgc gaaggatttg	540
	ccagggttcg ccaagatctc gcagcctgtat caaattacgc tgcttaaggc ttgtctcaagt	600
	gaggtaatga tgctccgagt cgcgcgcaga tacatgccc cttcagacag tggctgttc	660
20	gcgaaacaaac aaggctacac tccgcacaaac taccgcaagg ctggcatggc ctacgtcatc	720
	gaggatctac tgcaacttctg ccgggtcatg tactctatgg cggtggacaa catcattac	780
	gcgcgtgtca cggctgtcgt catctttct gaccggccag gggtggacca gcccactg	840
25	gtggaaagaaa tccagcggtt ctacatgtat acgctccgca tctatatactt gaaccagctg	900
	agcgggtcgg cgcgttcgtc cgtcatatac ggcgaatgtcc tctcaatctt ctctgagcta	960
	cgccacgtcg gcatgcaaaa ctccaaatcg tgcatctccc tcaagtcataa gaacagaaag	1020
30	ctggccgtt tccctcgagga gatctggat gtggcggacca tgtcgcacac ccaaccggccg	1080
	cctatcctcg agtccccccac gaatctctag	1110

&lt;210&gt; 3

&lt;211&gt; 1054

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Novel Sequence

&lt;400&gt; 3

45

50

55

210>4

<210> 4

<212> DNA

35 <213> Artificial Sequence

**<215> AIII**

<221> misc feature

<221> misc feature  
<222> Novel Sequence

-400- 4

taccaggacg ggtacgagca gccttctgtat gaagatttgaa agaggattac gcagacgtgg 60  
cagcaagccg acgataaaa cgaagacttgc gacactccct tccgcaggat cacagatgat 120  
actatctca cggtccaaact tatcgtggat ttgcgaaagg gattgccagg gttcccaaag 180  
atctcgacgc ctgatcaaat tacgctgtt aaggcttgc t caagtggat aatgtatgtc 240  
cgagtccccc gacgatacga tgcggctca gacagttgttc tggtccggaa caaccaaggc 300

	tacactcgcg acaactacctg caaggctggc atggcatacg tcatcgagga tctactgcac	360
5	ttctgcgggt gcatgtactc tatggcggtt gacaacatcc attacgcgt gtcacggct	420
	gtcgcatct tttctgaccg gccagggttg gagcagccgc aactgggta agaaatccag	480
	cggtaatacc tgatactatccatcattatcctgaaacc agctgagcgg gtcggcgct	540
10	tgcgtccgtca tatacggcaa gatccctctea atccctctg agctacgcac gtcggcatg	600
	caaaactcca acatgtgcat ctcccctaag ctcaagaaca gaaagctgcc gccttctctc	660
	gaggagatct gggatgtggc ggacatgtcg cacacccaac egccgcctat cctcgagtc	720
	cccacgaatac tcttag	735

15                   <210> 5  
                   <211> 960  
                   <212> DNA  
                   <213> Artificial Sequence  
 20                   <220>  
                   <221> misc\_feature  
                   <223> Novel Sequence  
                   <400> 5

25	ccttagtgcg tagtaaccga gactcagtgc gccatgaaac ggaaagagaa gaaagcacag	60
	aaggagaagg acaaactgcc tgtcgacacg acgacggtgg acgaccacat gccgcatt	120
	atgcagtgtg aaccctccacc tcctgaaagca gcaaggattc acgaagtgtt cccaggat	180
30	ctctccgaca agctgttgg acaaaaccgg cagaaaaaca tcccccaattt gacagccaa	240
	cagcagtcc ttatcgccag gtcatctgg taccaggacg ggtacgagca gccttctgtat	300
	gaagatttg aagaggattac gcagacgtgg cagcaagcgg acgtatggaaa cgaagagatct	360
35	gacactecct tccgcccattt cacagagatg actatctca cggtccaaact ttcgtggag	420
	ttcgcgaagg gattgccagg gttcgccaaat atctcgacgc ctgatccaaat tacgtcttt	480
	aaggcttgctt caagtggatg aatgtatgtc cggatcgccgac gacgatcgac tgccgcctca	540
40	gacagtgttc tgttcgccaa caaccatccgca tccatcgccg acaactaccg caaggctggc	600
	atggccatcg tcatcgagga tctactgcac ttctgggtt gcatgtactc tatggcggtt	660
	gacacatccatttacgcgt gtcacggct gtcgtcatct tttctgaccg gccagggttg	720
45	gagcagccgc aactgggta agaaatccag cggtactacc tgaatacgct ccgcatctat	780
	atccatcgacc agctgagcgg gtcggcgct tcgtccgtca tatacggcaa gatccctca	840
	atccctctgtg agctacgcac gtcggcatg caaaactcca acatgtgcat ctcccctaag	900
50	ctcaagaaca gaaagctgcc gccttctctc gaggagatct gggatgtggc ggacatgtcg	960

<210> 6  
   <211> 1878  
   <212> DNA  
 55                   <213> Artificial Sequence  
                   <220>  
                   <221> misc feature  
                   <223> Novel Sequence

&lt;400&gt; 6

	ggacctgcgc cacgggtgca agaggagctg tgccctggttt gccgcacag ggccctccggc	60
5	taccactaca acgcctcac ctgtgagggc tgcaaggggg tcttcgacg cagcgtaacg aaagcgccg tctactgctg caaattccgg cgccctcgca aaatggacat gtatcatgagg	120
	cgaaagtgtc aggagtgcgg cctgaaaaag tgccctggcc tgggtatgcg gccggaatgc	180
10	gtcgcccgg agaaccaatg tgcgatgaag cgccgcggaa agaaggccca gaaggagaag gacaaaatga ccacttcgcc gagcttcag catggggca atggcagctt ggcctctgg	240
	ggcgcccaag actttgttaa gaaggagatt cttgaccta tgacatcgca gccgccccag	300
15	catggacta ttccgctact acctgtatgaa atattggcca agtgtcaagc ggcgaatata ccttccttaa cgtacaatca gttggccgtt atatacaagt taatttgta ccaggatggc	360
	tatgagcagc catctgaaga ggatctcagg cgtataatga gtcacaccga tgagaacagag	420
20	agccaaacgg acgtcagctt tcgcatacccgatata acccgatata ccatactcac ggtccagg	480
	atttttgagt ttgcataagg tctaccagcg ttacaaaga taccccgga ggaccagatc	540
	acgttactaa aggccgtcgc tcggagggtt atatgtgcg gtatggcagc acgctatgac	600
25	cacagctcg actcaatatt ctccgcgaat aatacgatcat atacgcggga ttcttacaaa	660
	atggccggaa tggctgataa cattgaagac ctgcgcatt tctgcccggca aatgttctcg	720
	atgaagggtt acaacgtcgaa atacgcgtt ctcactgcga ttgtgatctt ctccggacccg	780
30	ccggccctgg agaaggccca actagtgcga cgcgttccaga gctactacat cgacacgcta	840
	cgccatttata tactcaaccg ccactgcggc gactcaatga gcctcgctt ctacgcggaa	900
	ctgctctega tcctcaccga gctgcgtacg ctgggcAAC agaacgcggaa gatgttttc	960
	tcaactaaacgc tcaaaaacccg caaactgcggc aagtctctgg agagatctg ggacgttcat	1020
35	gcacatcccgc categggttca gtcgcacccctt cagattaccc aggaggagaa cgacgcgttc	1080
	gagccggctg agcgtatgcg ggcgcgtt gggggccggca ttaccgcgg cattgtattgc	1140
	gactctgcctt ccacttcggc ggccgcggcc gcccggccggc atcagccctca gcctcgacccc	1200
40	cagccccaaac cctccctccctt gacccagaac gatccccccgc accagacaca gccgcagcta	1260
	caaccccgac taccacccca gctgcggatg caactgcggaa cccagctccca accacagctt	1320
	cagacccggaa tccaggccaca gatccaacca cagccacccgc tccctccctg ctccgcgtcc	1380
45	gtgccccctt ccgttaaccgc acctgggtcc ttgtccgcgg tcagtgatcg cagcgaatac	1440
	atggccggaa gtgcggccat aggaccatc acgcggccaa ccaccaggag tatcacggct	1500
	ggcgttacccg ctagctccac cacaatcgccg gtaccatgtt gcaacggagt tggagtccgg	1560
50		1620
	ggtgggggtgg gccggcaacgt cagcatgtat gcaacgcggc acacggcgat ggccttgatg	1680
	gggttagcccc tgcatttcgc ccaagagcagc ttatcgcccc gaggccgggt taagtccgg	1740
55	cactcgacga ctgcatacg	1800

&lt;210&gt; 7

<211> 1752  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 5 <221> misc feature  
 <223> Novel Sequence  
 <400> 7

10	gcccgtctact gctgcaagtt cgggcgcgcc tcgcgaatgg acatgtacat gaggcgaaag	60
	tgtcaggagt gcccgcgtaa aaagtgcgtg gccgtggta tgccgcgcga atgcgtcgct	120
	ccggagaacc aatgtgcgtat gaagccgcgc gaaaagaagg cccagaagga gaaggacaaa	180
15	atgaccacct cgccgagctc tcagcatggc ggcaatggca gcttggccctc tggccgcgc	240
	caagacttgc ttaagaagga gattcttgac cttatgacat gcgagccgc ccagcatgcc	300
	actattccgc tactacctgt tgaatattgc gccaagtgtc aagcgcgccaa tataccttcc	360
20	ttaacgtaca atcagtttgc cgttatatac aagttatattt ggttaccaggat tggctatgg	420
	cagccatctg aagaggatct caggcgatata atgactcaac ccgtatggaa cgagagccaa	480
	acggacgtca gctttccgc tataaccggat ataaccatata tcacggtcca gttgatttgtt	540
	gatgttgcata aaggttctacc aegcgtttaca aagatacccc aggaggacca gatcacgtta	600
25	ctaaaggcct gctcgccgc ggtgtatgt ctgcgtatgg cacgacgtca tgaccacacgc	660
	tccggactca tattttccgc gaataataga tcatatacgc gggatctta caaaatggcc	720
	ggaatggctg ataacattgt aagactgtcg cattttccgc gccaaatgtt ctgcgtatgg	780
30	gtggacaacgc tcgaataacgc gcttctact gccattgttgc ttttccgc cggccggc	840
	ctggagaagg cccaaactgt cgaaggcgtc cagacttact acatcgacac gctacgcatt	900
	tatatactca accgcctactg cggcgactca atgactctgc ttttctacgc aaagtcgtc	960
35	tccgtatca ccgagctgcg tacgtggc aaccagaacgc ccgagatgtg ttttctacta	1020
	aagctcaaaa accgcctact gcccaagtgc ttcgaggaga tctggacgt tcatgcctc	1080
	ccgcctatcg tccagtcgc ctttcgatcc acccaggagg agaacgcgcgctcgagccgg	1140
40	gctgagctga tgccgcgcgtc gtttggggcc gccattaccc ccggcatgtatggactt	1200
	gccttcactt ccggccgcgc accgcgcgcgc cagcatgcgc ctcagctca gccccagccc	1260
	caacccctctt ccctgaccca gaacgttcc cagcacccaga cacagccgcgactacaacact	1320
45	cagcttaccac ctcagctgcg aggtcaacttgc caaccccgaccc tccaaaccaca gcttcgacgc	1380

50

55

	caactccagc cacagattca accacagcca cagctccccc ccgtctccgc tccccgtgcc	1440
	gcctccgtaa ccgcacctgg ttcccttgcgc ggggtcagta cgagcagcga atacatgggc	1500
5	ggaagtgcgg ccataggacc catcacgcgg gcaaccacca gcgatatac ggctgcgcgtt	1560
	accgctagct ccaccacate a诶cggttaccc atgggcaacg gagtttgagg cggtgttggg	1620
	gtggccggca acgtcagcat gtatgcgaac gcccagacgg ctagggcctt gatgggtgt	1680
10	gcctgcatt cgaccaaga gcagttatc gggggagtgg cggttaagtc ggagcactcg	1740
	acgactgcat ag	1752

<210> 8  
 15 <211> 1650  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <221> misc feature  
 20 <223> Novel Sequence  
 <400> 8

25	cggccggaat gcgtcggtccc ggagaaccaa tggatgcgttga a诶cgccgcga aaagaaggcc	60
	cagaaggaga aggacaaaat gaccacttc ecgagatcttc a诶catggcgg caatggcagc	120
	ttggccctcg gtggccggca agactttgtt aagaaggaga ttcttgacat tatgacatgc	180
	gagccgcccc agcatgccc tattccgcta ctacctgtat aaatattggc caagtgtcaa	240
30	gcgcgaata taccttcctt aacgtacaat cagttggccg ttatatacaa gttaaatttttgg	300
	tacccaggat gctatgagca gccatctgaa gaggatctca ggcgtataat ggttcaacc	360
	gatggagaacg agagccaaac ggacgtcagc ttccggcata taaccggat aaccataactc	420
35	acggccatgt tgattgttga gtttgcataa gggttaccag cgtttacaaa gatccccag	480
	gaggaccaga tcacgttact aaaggcctgc tcgtccggagg tgatgtatgc gctgtatggca	540
	cgacgcctatg accacagctc ggactcaata ttcttcgcga ataataatgc atatacgccg	600
40	gatttttaca aaatggccgg aatggctgtat aacattgttgc acctgtcgca ttcttcgcgc	660
	caaatgttct cgatgttgcgg ggacaacgcgta gaatacgcc tccttcgtatc catttgtatc	720
	ttcttcggacc ggccggccgt ggagaaggcc caactgtcg aagcgatcca gagctactac	780
45	atcgacacgc tacgcattta tataactcaac cggccactgcg ggcactcaat gggccctgc	840
	ttctacgcggaa agctgtctc gatccctacc gagctgcgttgc cgcgtggccaa ccagaacgc	900
	gagatgttgc ttctactaaa gctaaaaac cggccactgc ccaagtcttgc gggggatc	960
50	tgggacgttc atgcattccc gccatcggtc cagtcgcacc ttccatggatttccaggag	1020
	aacgagcgtc tggggccggc tgagctgtatgc cggccatcggttgc cattaccgc	1080
	ggcattgatttgc ggcactctgc ctccactcg gcccggccag cccggcccca gcatcggct	1140
	caggctcagc cccagccccca acccttcctcc ctgaccggaga acgatccccca gcaccagaca	1200

	cagccgcage tacaacaccta gctaccaccc t cagctgc aag gtcaactgca accccc agetc	1260
	caaccac a cttc a gacgc a ctc c a gcca c agat cta ac cac a gcca a gtc ct tcc	1320
5	gtc tc cgg ctc c cg tgg ccc c ct cc gta acc gc ac ct gg tt ctt gtc cgc gtc a gtc a g	1380
	agc a gca gat a cat ggg cgg a a g tgc ggcc a ttag gaccc a tca cgc ggc a acc a cca c	1440
	a gta tca cgg ct g c gtt ac eg ct a gtc e acc a cat ca g cgg tacc gat ggg ca a cgg a	1500
10	gtt gg a gtc g gt gtt ggg ggt ggg cgg ca a c gtc a gcat g t atgc a a cgc c c a g a c g c	1560
	atgg cct tga tgg gtt gtc a cct gat tgc c a cca a g a g c a gtc tt a cgg ggg a gtc a g	1620
	gtta a gtc gg a gca ct gca gact gcat a g	1650

15  
<210> 9  
<211> 1338  
<212> DNA  
<213> Artificial Sequence  
20 <220>  
<221> misc feature  
<223> Novel Sequence  
<400> 9

25	tat gaggc a g c a t c t gaa ga g gat ctc a g g c tata a t g a g tca acccg a tgaga a c g a g	60
	agcc a a c a g c acgt c a g c t t cgg c a t a t a c c g a g a t a a c c t a c a c g t c a g t g	120
	at t gtt g a g t t g t g a g t t g t g a a a g g t c t a c c a g c g t t a c c a a a g a t a c c a g a g a t c	180
30	ac gtt tact a a a g g c t c t g c t c g c g a g g t g t a g t a g t c t g c g t a g g c a c g a c g t a t g a c	240
	c a c a g c t c g g a c t c a a t a t t c t t c g c g a a t a g a t c a t a t c g c g g a t t c t t a c a a a	300
	at g g c c g g a a t g g c t g t a a a t t c g a a g a c t g c t g c t t c t g c c a a a t g t t c t c g	360
35	at g a a g g t g g a c a a c g t c g a a t a c g e g c t t c t a c t g c c a t t g t g a t c t t c t g g a c c g g	420
	c c g g c c t c g g a g a a g g c c c a a t a g t c g a a g c g a t c c a g a t c a t c g a c a c g t a	480
	c g c a t t a t a a t a c t c a a c c g c a c t g e g g c g a c t c a a t g a g c t c t g t c t t c a c g c a a a g	540
40	c t g t c t c g a a t a c t c a c c g a g c t g c t a g c t g g c a a c c a g a a c c g c g a t g t g t t c	600
	t c a c t a a a a g c c a a a a c t g c c c a a a g t t c t c g a g g a g a t c t g g a c g t c t a t	660
	g c c a t c c c c g c a t c g g t c c a g t c g a c c t t c a g a t a c c c a g a g g a g a a c g c g t c t c	720
45	g a g c g g g c t g a g c t a t g c g g c t g g c a a c c a g a a c c g c g a t g t g t t c t c g c a a a g	780
	g a c t c t g c t c a c t t c g g c g c g g c a g c c g c g g c c a g c a t c a c g c t c a g c c c	840
	c a g c c c c a a c c t c t c c c t g a c c a g a a c g t t c c a g c a c a c a c g c a g c t a	900
50	c a a c t c a g c t a c t c a c c t c a g t c a a g g t a c a c t g c a a c c c a g c t c a a c c a g c t t	960
	c a g a c g c a c a t c c a g c c a c a g t t c a a c c a g c a c a g c t c t c c g t c t c c	1020
	g t g c c c g c t t c g t a a c c g c a c t g g t t c c t g t c c g c g g c t a g t a c g a g a t a c	1080

55

	atgggcggaa gtgcggccat aggaccate acgcccga caaccacgat tatcacggct	1140
	gccgttaccc cttagtcccc cacatcagcg gtaccatgg gcaacggagt tggactcggt	1200
5	gttgggggtt gcccccaacgt cagcatgtat gccaacgccc agacggcgat ggccttgatg	1260
	ggtgttagccc tgcatttgcga ccaagagcag cttatcgaaa gaggcccggt taagtccggag	1320
	cactcgacga ctgcatacg	1338

10

<210> 10  
<211> 969  
<212> DNA  
<213> Artificial Sequence  
15 <220>  
<221> misc feature  
<223> Novel Sequence  
<400> 10

20	cggccggaaat gcgtcgatccc ggagaaccaa ttgtcgatga agccggcgaa aagaaggcc	60
	cagaaggaga aggacaaaat gaccatctcg ccgagcttc agcatggcg caatggcagc	120
	ttggccctcg gtggcgccca agactttgtt aagaaggaga ttcttgaccc tatgacatgc	180
25	gaggccccc agcatgccc tattccgtta ctacccatgat aaatattggc caagtgtcaa	240
	gcgcgcataa taccccttcc aacgtacaat cagttggcg ttatatacaa gttaaatttgg	300
	taccaggatg gctatgacca gccatctgaa gaggatctca ggcgtataat gagtcaaccc	360
30	gatgagaacg agagccaaac ggacgtcage ttccggata taaccgagat aaccataactc	420
	acggtccagt tgatgttga gtttgcataa ggttctaccag ctgttaccaa gataccccag	480
	gaggaccaga tcacgttact aaaggccctgc tcgtccgggg tgatgtatgtc gctgtatggca	540
35	cgacgcctatg accacagctc ggactcaata ttcttcgca ataatagatc atatacgccg	600
	gattcttaca aaatggcccg aatggctgtat aacattgaag acctgtcgtca ttcttcggcc	660
	caaatgttct cgatgaaagg .ggacaacgtc gaatacgccgc ttctcaactgc cattgtgtatc	720
40	ttcttcggacc ggccggccct ggagaaggcc caactagtgc aagcgatcca gagctactac	780
	atcgacacgc tacgcattta tataactcaac cgccactgcg ggcactcaat ggcctcgatc	840
	ttcttcggccaa agctgtctc gatccatccacc gagctcgatca cgctggccaa ccagaacgccc	900
45	gagatgtgtt tctcaactaaa gctcaaaaac cgccaaactgc ccaaggccct cgaggagatc	960
	tgggacgtt	969

<210> 11  
<211> 412  
50 <212> PRT  
<213> Artificial Sequence  
<220>  
<221> misc feature  
<223> Novel Sequence  
55 <400> 11

Lys Gly Pro Ala Pro Arg Gln Gln Glu Leu Cys Leu Val Cys Gly  
 1 5 10 15  
 5 Asp Arg Ala Ser Gly Tyr His Tyr Asn Ala Leu Thr Cys Glu Gly Cys  
 20 25 30  
 Lys Gly Phe Phe Arg Arg Ser Val Thr Lys Asn Ala Val Tyr Ile Cys  
 35 40 45  
 10 Lys Phe Gly His Ala Cys Glu Met Asp Met Tyr Met Arg Arg Lys Cys  
 50 55 60  
 Gln Glu Cys Arg Leu Lys Lys Cys Leu Ala Val Gly Met Arg Pro Glu  
 65 70 75 80  
 15 Cys Val Val Pro Glu Thr Gln Cys Ala Met Lys Arg Lys Glu Lys Lys  
 85 90 95  
 Ala Gln Lys Glu Lys Asp Lys Leu Pro Val Ser Thr Thr Val Asp  
 100 105 110  
 20 Asp His Met Pro Pro Ile Met Gln Cys Glu Pro Pro Pro Glu Ala  
 115 120 125  
 Ala Arg Ile His Glu Val Val Pro Arg Phe Leu Ser Asp Lys Leu Leu  
 130 135 140  
 25 Glu Thr Asn Arg Gln Lys Asn Ile Pro Gln Leu Thr Ala Asn Gln Gln  
 145 150 155 160  
 Phe Leu Ile Ala Arg Leu Ile Trp Tyr Gln Asp Gly Tyr Glu Gln Pro  
 165 170 175  
 30 Ser Asp Glu Asp Leu Lys Arg Ile Thr Gln Thr Trp Gln Gln Ala Asp  
 180 185 190  
 Asp Glu Asn Glu Glu Ser Asp Thr Pro Phe Arg Gln Ile Thr Glu Met  
 195 200 205  
 Thr Ile Leu Thr Val Gln Leu Ile Val Glu Phe Ala Lys Gly Leu Pro  
 210 215 220  
 35 Gly Phe Ala Lys Ile Ser Gln Pro Asp Gln Ile Thr Leu Leu Lys Ala  
 225 230 235 240  
 Cys Ser Ser Glu Val Met Met Leu Arg Val Ala Arg Arg Tyr Asp Ala  
 245 250 255  
 40 Ala Ser Asp Ser Val Leu Phe Ala Asn Asn Gln Ala Tyr Thr Arg Asp  
 260 265 270  
 Asn Tyr Arg Lys Ala Gly Met Ala Tyr Val Ile Glu Asp Leu Leu His  
 275 280 285  
 45 Phe Cys Arg Cys Met Tyr Ser Met Ala Leu Asp Asn Ile His Tyr Ala  
 290 295 300  
 Leu Leu Thr Ala Val Val Ile Phe Ser Asp Arg Pro Gly Leu Glu Gln  
 305 310 315 320  
 50 Pro Gln Leu Val Glu Glu Ile Gln Arg Tyr Tyr Leu Asn Thr Leu Arg



Asp Glu Asn Glu Glu Ser Asp Thr Pro Phe Arg Gln Ile Thr Glu Met  
 195 200 205  
 5 Thr Ile Leu Thr Val Gln Leu Ile Val Glu Phe Ala Lys Gly Leu Pro  
 210 215 220  
 Gly Phe Ala Lys Ile Ser Gln Pro Asp Gln Ile Thr Leu Leu Lys Ala  
 225 230 235 240  
 10 Cys Ser Ser Glu Val Met Met Leu Arg Val Ala Arg Arg Tyr Asp Ala  
 245 250 255  
 Ala Ser Asp Ser Val Leu Phe Ala Asn Asn Gln Ala Tyr Thr Arg Asp  
 260 265 270  
 Asn Tyr Arg Lys Ala Gly Met Ala Tyr Val Ile Glu Asp Leu Leu His  
 15 275 280 285  
 Phe Cys Arg Cys Met Tyr Ser Met Ala Leu Asp Asn Ile His Tyr Ala  
 290 295 300  
 20 Leu Leu Thr Ala Val Val Ile Phe Ser Asp Arg Pro Gly Leu Glu Gln  
 305 310 315 320  
 Pro Gln Leu Val Glu Glu Ile Gln Arg Tyr Tyr Leu Asn Thr Leu Arg  
 325 330 335  
 Ile Tyr Ile Leu Asn Gln Leu Ser Gly Ser Ala Arg Ser Ser Val Ile  
 340 345 350  
 25 Tyr Gly Lys Ile Leu Ser Ile Leu Ser Glu Leu Arg Thr Leu Gly Met  
 355 360 365  
 Gln Asn Ser Asn Met Cys Ile Ser Leu Lys Leu Lys Asn Arg Lys Leu  
 370 375 380  
 30 Pro Pro Phe Leu Glu Ile Trp Asp Val Ala Asp Met Ser His Thr  
 385 390 395 400  
 Gln Pro Pro Pro Ile Leu Glu Ser Pro Thr Asn Leu  
 405 410  
 35 <210> 13  
 <211> 334  
 <212> PRT  
 <213> Artificial Sequence  
 40 <220>  
 <221> misc feature  
 <223> Novel Sequence  
 <400> 13

45 Pro Glu Cys Val Val Pro Glu Thr Gln Cys Ala Met Lys Arg Lys Glu  
 1 5 10 15  
 Lys Lys Ala Gln Lys Glu Lys Asp Lys Leu Pro Val Ser Thr Thr  
 20 25 30  
 50 Val Asp Asp His Met Pro Pro Ile Met Gln Cys Glu Pro Pro Pro Pro  
 35 40 45  
 Glu Ala Ala Arg Ile His Glu Val Val Pro Arg Phe Leu Ser Asp Lys  
 50 55 60

Leu Leu Glu Thr Asn Arg Gln Lys Asn Ile Pro Gln Leu Thr Ala Asn  
 65 70 75 80

5 Gln Gln Phe Leu Ile Ala Arg Leu Ile Trp Tyr Gln Asp Gly Tyr Glu  
 85 90 95

Gln Pro Ser Asp Glu Asp Leu Lys Arg Ile Thr Gln Thr Trp Gln Gln  
 100 105 110

10 Ala Asp Asp Glu Asn Glu Glu Ser Asp Thr Pro Phe Arg Gln Ile Thr  
 115 120 125

Glu Met Thr Ile Leu Thr Val Gln Leu Ile Val Glu Phe Ala Lys Gly  
 130 135 140

15 Leu Pro Gly Phe Ala Lys Ile Ser Gln Pro Asp Gln Ile Thr Leu Leu  
 145 150 155 160

Lys Ala Cys Ser Ser Glu Val Met Met Leu Arg Val Ala Arg Arg Tyr  
 165 170 175

20 Asp Ala Ala Ser Asp Ser Val Leu Phe Ala Asn Asn Gln Ala Tyr Thr  
 180 185 190

Arg Asp Asn Tyr Arg Lys Ala Gly Met Ala Tyr Val Ile Glu Asp Leu  
 195 200 205

25 Leu His Phe Cys Arg Cys Met Tyr Ser Met Ala Leu Asp Asn Ile His  
 210 215 220

Tyr Ala Leu Leu Thr Ala Val Val Ile Phe Ser Asp Arg Pro Gly Leu  
 225 230 235 240

30 Glu Gln Pro Gln Leu Val Glu Glu Ile Gln Arg Tyr Tyr Leu Asn Thr  
 245 250 255

Leu Arg Ile Tyr Ile Leu Asn Gln Leu Ser Gly Ser Ala Arg Ser Ser  
 260 265 270

35 Val Ile Tyr Gly Lys Ile Leu Ser Ile Leu Ser Glu Leu Arg Thr Leu  
 275 280 285

Gly Met Gln Asn Ser Asn Met Cys Ile Ser Leu Lys Leu Lys Asn Arg  
 290 295 300

Lys Leu Pro Pro Phe Leu Glu Glu Ile Trp Asp Val Ala Asp Met Ser  
 305 310 315 320

40 His Thr Gln Pro Pro Pro Ile Leu Glu Ser Pro Thr Asn Leu  
 325 330

&lt;210&gt; 14

&lt;211&gt; 244

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;221&gt; misc feature

&lt;223&gt; Novel Sequence

&lt;400&gt; 14

Tyr Gln Asp Gly Tyr Glu Gln Pro Ser Asp Glu Asp Leu Lys Arg Ile

1	5	10	15
	Thr Gln Thr Trp Gln Gln Ala Asp Asp Glu Asn Glu Glu Ser Asp Thr		
	20 25 30	25	30
5	Pro Phe Arg Gln Ile Thr Glu Met Thr Ile Leu Thr Val Gln Leu Ile		
	35 40 45	40	45
	Val Glu Phe Ala Lys Gly Leu Pro Gly Phe Ala Lys Ile Ser Gln Pro		
	50 55 60	55	60
10	Asp Gln Ile Thr Leu Leu Lys Ala Cys Ser Ser Glu Val Met Met Leu		
	65 70 75 80	70	75 80
	Arg Val Ala Arg Arg Tyr Asp Ala Ala Ser Asp Ser Val Leu Phe Ala		
	85 90 95	85	90 95
15	Asn Asn Gln Ala Tyr Thr Arg Asp Asn Tyr Arg Lys Ala Gly Met Ala		
	100 105 110	105	110
	Tyr Val Ile Glu Asp Leu Leu His Phe Cys Arg Cys Met Tyr Ser Met		
	115 120 125	120	125
20	Ala Leu Asp Asn Ile His Tyr Ala Leu Leu Thr Ala Val Val Ile Phe		
	130 135 140	135	140
	Ser Asp Arg Pro Gly Leu Glu Gln Pro Gln Leu Val Glu Glu Ile Gln		
	145 150 155 160	150	155 160
25	Arg Tyr Tyr Leu Asn Thr Leu Arg Ile Tyr Ile Leu Asn Gln Leu Ser		
	165 170 175	165	170 175
	Gly Ser Ala Arg Ser Ser Val Ile Tyr Gly Lys Ile Leu Ser Ile Leu		
	180 185 190	180	185 190
30	Ser Glu Leu Arg Thr Leu Gly Met Gln Asn Ser Asn Met Cys Ile Ser		
	195 200 205	195	200 205
	Leu Lys Leu Lys Asn Arg Lys Leu Pro Pro Phe Leu Glu Glu Ile Trp		
	210 215 220	215	220
35	Asp Val Ala Asp Met Ser His Thr Gln Pro Pro Pro Ile Leu Glu Ser		
	225 230 235 240	225	230 235 240
	Pro Thr Asn Leu		

40 <210> 15  
 <211> 320  
 <212> PRT  
 <213> Artificial Sequence  
 <220>  
 45 <221> misc feature  
 <223> Novel Sequence  
 <400> 15

50	Pro Glu Cys Val Val Pro Glu Thr Gln Cys Ala Met Lys Arg Lys Glu		
	1 5 10 15	5	10 15
	Lys Lys Ala Gln Lys Glu Lys Asp Lys Leu Pro Val Ser Thr Thr Thr		
	20 25 30	20	25 30

Val Asp Asp His Met Pro Pro Ile Met Gln Cys Glu Pro Pro Pro Pro  
 35 40 45  
 5 Glu Ala Ala Arg Ile His Glu Val Val Pro Arg Phe Leu Ser Asp Lys  
 50 55 60  
 Leu Leu Glu Thr Asn Arg Gln Lys Asn Ile Pro Gln Leu Thr Ala Asn  
 65 70 75 80  
 10 Gln Gln Phe Leu Ile Ala Arg Leu Ile Trp Tyr Gln Asp Gly Tyr Glu  
 85 90 95  
 Gln Pro Ser Asp Glu Asp Leu Lys Arg Ile Thr Gln Thr Trp Gln Gln  
 100 105 110  
 15 Ala Asp Asp Glu Asn Glu Glu Ser Asp Thr Pro Phe Arg Gln Ile Thr  
 115 120 125  
 Glu Met Thr Ile Leu Thr Val Gln Leu Ile Val Glu Phe Ala Lys Gly  
 130 135 140  
 20 Leu Pro Gly Phe Ala Lys Ile Ser Gln Pro Asp Gln Ile Thr Leu Leu  
 145 150 155 160  
 Lys Ala Cys Ser Ser Glu Val Met Met Leu Arg Val Ala Arg Arg Tyr  
 165 170 175  
 Asp Ala Ala Ser Asp Ser Val Leu Phe Ala Asn Asn Gln Ala Tyr Thr  
 180 185 190  
 25 Arg Asp Asn Tyr Arg Lys Ala Gly Met Ala Tyr Val Ile Glu Asp Leu  
 195 200 205  
 Leu His Phe Cys Arg Cys Met Tyr Ser Met Ala Leu Asp Asn Ile His  
 210 215 220  
 30 Tyr Ala Leu Leu Thr Ala Val Val Ile Phe Ser Asp Arg Pro Gly Leu  
 225 230 235 240  
 Glu Gln Pro Gln Leu Val Glu Glu Ile Gln Arg Tyr Tyr Leu Asn Thr  
 245 250 255  
 35 Leu Arg Ile Tyr Ile Leu Asn Gln Leu Ser Gly Ser Ala Arg Ser Ser  
 260 265 270  
 Val Ile Tyr Gly Lys Ile Leu Ser Ile Leu Ser Glu Leu Arg Thr Leu  
 275 280 285  
 40 Gly Met Gln Asn Ser Asn Met Cys Ile Ser Leu Lys Leu Lys Asn Arg  
 290 295 300  
 Lys Leu Pro Pro Phe Leu Glu Ile Trp Asp Val Ala Asp Met Ser  
 305 310 315 320

45 <210> 16  
 <211> 625  
 <212> PRT  
 <213> Artificial Sequence  
 50 <220>  
 <221> misc feature  
 <223> Novel Sequence  
 <400> 16

Gly Pro Ala Pro Arg Val Gln Glu Glu Leu Cys Leu Val Cys Gly Asp  
 1 5 10 15  
 Arg Ala Ser Gly Tyr His Tyr Asn Ala Leu Thr Cys Glu Gly Cys Lys  
 5 20 25 30  
 Gly Phe Phe Arg Arg Ser Val Thr Lys Ser Ala Val Tyr Cys Cys Lys  
 35 40 45  
 Phe Gly Arg Ala Cys Glu Met Asp Met Tyr Met Arg Arg Lys Cys Gln  
 10 50 55 60  
 Glu Cys Arg Leu Lys Lys Cys Leu Ala Val Gly Met Arg Pro Glu Cys  
 15 65 70 75 80  
 Val Val Pro Glu Asn Gln Cys Ala Met Lys Arg Arg Glu Lys Lys Ala  
 85 90 95  
 Gln Lys Glu Lys Asp Lys Met Thr Thr Ser Pro Ser Ser Gln His Gly  
 100 105 110  
 Gly Asn Gly Ser Leu Ala Ser Gly Gly Gln Asp Phe Val Lys Lys  
 20 115 120 125  
 Glu Ile Leu Asp Leu Met Thr Cys Glu Pro Pro Gln His Ala Thr Ile  
 130 135 140  
 Pro Leu Leu Pro Asp Glu Ile Leu Ala Lys Cys Gln Ala Arg Asn Ile  
 25 145 150 155 160  
 Pro Ser Leu Thr Tyr Asn Gln Leu Ala Val Ile Tyr Lys Leu Ile Trp  
 165 170 175  
 Tyr Gln Asp Gly Tyr Glu Gln Pro Ser Glu Glu Asp Leu Arg Arg Ile  
 30 180 185 190  
 Met Ser Gln Pro Asp Glu Asn Glu Ser Gln Thr Asp Val Ser Phe Arg  
 195 200 205  
 His Ile Thr Glu Ile Thr Ile Leu Thr Val Gln Leu Ile Val Glu Phe  
 210 215 220  
 Ala Lys Gly Leu Pro Ala Phe Thr Lys Ile Pro Gln Glu Asp Gln Ile  
 35 225 230 235 240  
 Thr Leu Leu Lys Ala Cys Ser Ser Glu Val Met Met Leu Arg Met Ala  
 245 250 255  
 Arg Arg Tyr Asp His Ser Ser Asp Ser Ile Phe Phe Ala Asn Asn Arg  
 40 260 265 270  
 Ser Tyr Thr Arg Asp Ser Tyr Lys Met Ala Gly Met Ala Asp Asn Ile  
 275 280 285  
 Glu Asp Leu Leu His Phe Cys Arg Gln Met Phe Ser Met Lys Val Asp  
 45 290 295 300  
 Asn Val Glu Tyr Ala Leu Leu Thr Ala Ile Val Ile Phe Ser Asp Arg  
 305 310 315 320  
 Pro Gly Leu Glu Lys Ala Gln Leu Val Glu Ala Ile Gln Ser Tyr Tyr  
 50 325 330 335  
 Ile Asp Thr Leu Arg Ile Tyr Ile Leu Asn Arg His Cys Gly Asp Ser

	340	345	350	
	Met Ser Leu Val Phe Tyr Ala Lys	Leu Leu Ser Ile Leu Thr Glu Leu		
	355	360	365	
5	Arg Thr Leu Gly Asn Gin Asn Ala Glu Met Cys	Phe Ser Leu Lys Leu		
	370	375	380	
	Lys Asn Arg Lys Leu Pro Lys	Phe Leu Glu Glu Ile Trp Asp Val His		
	385	390	395	400
10	Ala Ile Pro Pro Val Gln Ser His	Leu Gln Ile Thr Gln Glu Glu		
	405	410	415	
	Asn Glu Arg Leu Glu Arg Ala Glu Arg Met Arg Ala Ser Val Gly Gly			
	420	425	430	
15	Ala Ile Thr Ala Gly Ile Asp Cys Asp Ser Ala Ser Thr Ser Ala Ala			
	435	440	445	
	Ala Ala Ala Ala Gln His Gln Pro Gln Pro Gln Pro Gln Pro			
	450	455	460	
20	Ser Ser Leu Thr Gln Asn Asp Ser Gln His Gln Thr Gln Pro Gln Leu			
	465	470	475	480
	Gln Pro Gln Leu Pro Pro Gln Leu Gln Gly Gln Leu Gln Pro Gln Leu			
	485	490	495	
25	Gln Pro Gln Leu Gln Thr Gln Leu Gln Pro Gln Ile Gln Pro Gln Pro			
	500	505	510	
	Gln Leu Leu Pro Val Ser Ala Pro Val Pro Ala Ser Val Thr Ala Pro			
	515	520	525	
30	Gly Ser Leu Ser Ala Val Ser Thr Ser Ser Glu Tyr Met Gly Gly Ser			
	530	535	540	
	Ala Ala Ile Gly Pro Ile Thr Pro Ala Thr Thr Ser Ser Ile Thr Ala			
	545	550	555	560
35	Ala Val Thr Ala Ser Ser Thr Thr Ser Ala Val Pro Met Gly Asn Gly			
	565	570	575	
	Val Gly Val Gly Val Gly Gly Asn Val Ser Met Tyr Ala Asn			
	580	585	590	
40	Ala Gln Thr Ala Met Ala Leu Met Gly Val Ala Leu His Ser His Gln			
	595	600	605	
	Glu Gln Leu Ile Gly Gly Val Ala Val Lys Ser Glu His Ser Thr Thr			
	610	615	620	
45	Ala			
	625			

&lt;210&gt; 17

&lt;211&gt; 583

&lt;212&gt; PRT

50 &lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;221&gt; misc feature

&lt;223&gt; Novel Sequence

&lt;400&gt; 17

55

Ala Val Tyr Cys Cys Lys Phe Gly Arg Ala Cys Glu Met Asp Met Tyr  
 1 5 10 15  
 Met Arg Arg Lys Cys Gln Glu Cys Arg Leu Lys Lys Cys Leu Ala Val  
 5 20 25 30  
 Gly Met Arg Pro Glu Cys Val Val Pro Glu Asn Gln Cys Ala Met Lys  
 10 35 40 45  
 Arg Arg Glu Lys Lys Ala Gln Lys Glu Lys Asp Lys Met Thr Thr Ser  
 50 55 60  
 Pro Ser Ser Gln His Gly Gly Asn Gly Ser Leu Ala Ser Gly Gly  
 65 70 75 80  
 Gln Asp Phe Val Lys Lys Glu Ile Leu Asp Leu Met Thr Cys Glu Pro  
 15 85 90 95  
 Pro Gln His Ala Thr Ile Pro Leu Leu Pro Asp Glu Ile Leu Ala Lys  
 100 105 110  
 Cys Gln Ala Arg Asn Ile Pro Ser Leu Thr Tyr Asn Gln Leu Ala Val  
 20 115 120 125  
 Ile Tyr Lys Leu Ile Trp Tyr Gln Asp Gly Tyr Glu Gln Pro Ser Glu  
 130 135 140  
 Glu Asp Leu Arg Arg Ile Met Ser Gln Pro Asp Glu Asn Glu Ser Gln  
 25 145 150 155 160  
 Thr Asp Val Ser Phe Arg His Ile Thr Glu Ile Thr Ile Leu Thr Val  
 165 170 175  
 Gln Leu Ile Val Glu Phe Ala Lys Gly Leu Pro Ala Phe Thr Lys Ile  
 30 180 185 190  
 Pro Gln Glu Asp Gln Ile Thr Leu Leu Lys Ala Cys Ser Ser Glu Val  
 195 200 205  
 Met Met Leu Arg Met Ala Arg Arg Tyr Asp His Ser Ser Asp Ser Ile  
 210 215 220  
 Phe Phe Ala Asn Asn Arg Ser Tyr Thr Arg Asp Ser Tyr Lys Met Ala  
 35 225 230 235 240  
 Gly Met Ala Asp Asn Ile Glu Asp Leu Leu His Phe Cys Arg Gln Met  
 245 250 255  
 Phe Ser Met Lys Val Asp Asn Val Glu Tyr Ala Leu Leu Thr Ala Ile  
 40 260 265 270  
 Val Ile Phe Ser Asp Arg Pro Gly Leu Glu Lys Ala Gln Leu Val Glu  
 275 280 285  
 Ala Ile Gln Ser Tyr Tyr Ile Asp Thr Leu Arg Ile Tyr Ile Leu Asn  
 45 290 295 300  
 Arg His Cys Gly Asp Ser Met Ser Leu Val Phe Tyr Ala Lys Leu Leu  
 305 310 315 320  
 Ser Ile Leu Thr Glu Leu Arg Thr Leu Gly Asn Gln Asn Ala Glu Met  
 50 325 330 335

Cys Phe Ser Leu Lys Leu Lys Asn Arg Lys Leu Pro Lys Phe Leu Glu  
 340 345 350

5 Glu Ile Trp Asp Val His Ala Ile Pro Pro Ser Val Gln Ser His Leu  
 355 360 365

Gln Ile Thr Gln Glu Glu Asn Glu Arg Leu Glu Arg Ala Glu Arg Met  
 370 375 380

10 Arg Ala Ser Val Gly Gly Ala Ile Thr Ala Gly Ile Asp Cys Asp Ser  
 385 390 395 400

Ala Ser Thr Ser Ala Ala Ala Ala Ala Gln His Gln Pro Gln Pro  
 405 410 415

15 Gln Pro Gln Pro Gln Pro Ser Ser Leu Thr Gln Asn Asp Ser Gln His  
 420 425 430

Gln Thr Gln Pro Gln Leu Gln Pro Gln Leu Pro Pro Gln Leu Gln Gly  
 435 440 445

20 Gln Leu Gln Pro Gln Leu Gln Pro Gln Leu Gln Thr Gln Leu Gln Pro  
 450 455 460

Gln Ile Gln Pro Gln Pro Gln Leu Leu Pro Val Ser Ala Pro Val Pro  
 465 470 475 480

25 Ala Ser Val Thr Ala Pro Gly Ser Leu Ser Ala Val Ser Thr Ser Ser  
 485 490 495

Glu Tyr Met Gly Gly Ser Ala Ala Ile Gly Pro Ile Thr Pro Ala Thr  
 500 505 510

30 Thr Ser Ser Ile Thr Ala Ala Val Thr Ala Ser Ser Thr Thr Ser Ala  
 515 520 525

Val Pro Met Gly Asn Gly Val Gly Val Gly Val Gly Gly Asn  
 530 535 540

35 Val Ser Met Tyr Ala Asn Ala Gln Thr Ala Met Ala Leu Met Gly Val  
 545 550 555 560

Ala Leu His Ser His Gln Glu Gln Leu Ile Gly Gly Val Ala Val Lys  
 565 570 575

40 Ser Glu His Ser Thr Thr Ala  
 580

45 <210> 18

<211> 549

<212> PRT

<213> Artificial Sequence

<400> 18

Arg Pro Glu Cys Val Val Pro Glu Asn Gln Cys Ala Met Lys Arg Arg  
 1 5 10 15

50 Glu Lys Lys Ala Gln Lys Glu Lys Asp Lys Met Thr Thr Ser Pro Ser  
 20 25 30

Ser Gln His Gly Gly Asn Gly Ser Leu Ala Ser Gly Gly Gln Asp  
 35 40 45

Phe Val Lys Lys Glu Ile Leu Asp Leu Met Thr Cys Glu Pro Pro Gln  
 50 55 . 60  
 His Ala Thr Ile Pro Leu Leu Pro Asp Glu Ile Leu Ala Lys Cys Gln  
 65 70 75 80  
 Ala Arg Asn Ile Pro Ser Leu Thr Tyr Asn Gln Leu Ala Val Ile Tyr  
 85 90 95  
 Lys Leu Ile Trp Tyr Gln Asp Gly Tyr Glu Gln Pro Ser Glu Glu Asp  
 100 105 110  
 Leu Arg Arg Ile Met Ser Gln Pro Asp Glu Asn Glu Ser Gln Thr Asp  
 115 120 125  
 Val Ser Phe Arg His Ile Thr Glu Ile Thr Ile Leu Thr Val Gln Leu  
 130 135 140  
 Ile Val Glu Phe Ala Lys Gly Leu Pro Ala Phe Thr Lys Ile Pro Gln  
 145 150 155 160  
 Glu Asp Gln Ile Thr Leu Leu Lys Ala Cys Ser Ser Glu Val Met Met  
 165 170 175  
 Leu Arg Met Ala Arg Arg Tyr Asp His Ser Ser Asp Ser Ile Phe Phe  
 180 185 190  
 Ala Asn Asn Arg Ser Tyr Thr Arg Asp Ser Tyr Lys Met Ala Gly Met  
 195 200 205  
 Ala Asp Asn Ile Glu Asp Leu Leu His Phe Cys Arg Gln Met Phe Ser  
 210 215 220  
 Met Lys Val Asp Asn Val Glu Tyr Ala Leu Leu Thr Ala Ile Val Ile  
 225 230 235 240  
 Phe Ser Asp Arg Pro Gly Leu Glu Lys Ala Gln Leu Val Glu Ala Ile  
 245 250 255  
 Gln Ser Tyr Tyr Ile Asp Thr Leu Arg Ile Tyr Ile Leu Asn Arg His  
 260 265 270  
 Cys Gly Asp Ser Met Ser Leu Val Phe Tyr Ala Lys Leu Leu Ser Ile  
 275 280 285  
 Leu Thr Glu Leu Arg Thr Leu Gly Asn Gln Asn Ala Glu Met Cys Phe  
 290 295 300  
 Ser Leu Lys Leu Lys Asn Arg Lys Leu Pro Lys Phe Leu Glu Glu Ile  
 305 310 315 320  
 Trp Asp Val His Ala Ile Pro Pro Ser Val Gln Ser His Leu Gln Ile  
 325 330 335  
 Thr Gln Glu Glu Asn Glu Arg Leu Glu Arg Ala Glu Arg Met Arg Ala  
 340 345 350  
 Ser Val Gly Gly Ala Ile Thr Ala Gly Ile Asp Cys Asp Ser Ala Ser  
 355 360 365  
 Thr Ser Ala Ala Ala Ala Ala Gln His Gln Pro Gln Pro Gln Pro  
 370 375 380  
 Gln Pro Gln Pro Ser Ser Leu Thr Gln Asn Asp Ser Gln His Gln Thr  
 385 390 395 400

Gln Pro Gln Leu Gln Pro Gln Leu Pro Pro Gln Leu Gln Gly Gln Leu  
 405 410 415  
 5 Gln Pro Gln Leu Gln Pro Gln Leu Gln Thr Gln Leu Gln Pro Gln Ile  
 420 425 430  
 Gln Pro Gln Pro Gln Leu Leu Pro Val Ser Ala Pro Val Pro Ala Ser  
 435 440 445  
 10 Val Thr Ala Pro Gly Ser Leu Ser Ala Val Ser Thr Ser Ser Glu Tyr  
 450 455 460  
 Met Gly Gly Ser Ala Ala Ile Gly Pro Ile Thr Pro Ala Thr Thr Ser  
 465 470 475 480  
 15 Ser Ile Thr Ala Ala Val Thr Ala Ser Ser Thr Thr Ser Ala Val Pro  
 485 490 495  
 Met Gly Asn Gly Val Gly Val Gly Val Gly Gly Asn Val Ser  
 500 505 510  
 20 Met Tyr Ala Asn Ala Gln Thr Ala Met Ala Leu Met Gly Val Ala Leu  
 515 520 525  
 His Ser His Gln Glu Gln Leu Ile Gly Gly Val Ala Val Lys Ser Glu  
 530 535 540  
 25 His Ser Thr Thr Ala  
 545

<210> 19  
 <211> 445  
 <212> PRT  
 30 <213> Artificial Sequence  
 <400> 19

35 Tyr Glu Gln Pro Ser Glu Glu Asp Leu Arg Arg Ile Met Ser Gln Pro  
 1 5 10 15  
 Asp Glu Asn Glu Ser Gln Thr Asp Val Ser Phe Arg His Ile Thr Glu  
 20 25 30  
 40 Ile Thr Ile Leu Thr Val Gln Leu Ile Val Glu Phe Ala Lys Gly Leu  
 35 40 45  
 Pro Ala Phe Thr Lys Ile Pro Gln Glu Asp Gln Ile Thr Leu Leu Lys  
 50 55 60  
 45 Ala Cys Ser Ser Glu Val Met Met Leu Arg Met Ala Arg Arg Tyr Asp  
 65 70 75 80  
 His Ser Ser Asp Ser Ile Phe Phe Ala Asn Asn Arg Ser Tyr Thr Arg  
 85 90 95  
 50 Asp Ser Tyr Lys Met Ala Gly Met Ala Asp Asn Ile Glu Asp Leu Leu  
 100 105 110  
 His Phe Cys Arg Gln Met Phe Ser Met Lys Val Asp Asn Val Glu Tyr  
 115 120 125  
 Ala Leu Leu Thr Ala Ile Val Ile Phe Ser Asp Arg Pro Gly Leu Glu  
 130 135 140

Lys Ala Gln Leu Val Glu Ala Ile Gln Ser Tyr Tyr Ile Asp Thr Leu  
 145 150 155 160

Arg Ile Tyr Ile Leu Asn Arg His Cys Gly Asp Ser Met Ser Leu Val  
 165 170 175

5 Phe Tyr Ala Lys Leu Leu Ser Ile Leu Thr Glu Leu Arg Thr Leu Gly  
 180 185 190

Asn Gln Asn Ala Glu Met Cys Phe Ser Leu Lys Leu Lys Asn Arg Lys  
 195 200 205

10 Leu Pro Lys Phe Leu Glu Glu Ile Trp Asp Val His Ala Ile Pro Pro  
 210 215 220

Ser Val Gln Ser His Leu Gln Ile Thr Gln Glu Glu Asn Glu Arg Leu  
 225 230 235 240

15 Glu Arg Ala Glu Arg Met Arg Ala Ser Val Gly Gly Ala Ile Thr Ala  
 245 250 255

Gly Ile Asp Cys Asp Ser Ala Ser Thr Ser Ala Ala Ala Ala Ala  
 260 265 270

20 Gln His Gln Pro Gln Pro Gln Pro Gln Pro Ser Ser Leu Thr  
 275 280 285

Gln Asn Asp Ser Gln His Gln Thr Gln Pro Gln Leu Gln Pro Gln Leu  
 290 295 300

25 Pro Pro Gln Leu Gln Gly Gln Leu Gln Pro Gln Leu Gln Pro Gln Leu  
 305 310 315 320

Gln Thr Gln Leu Gln Pro Gln Ile Gln Pro Gln Pro Gln Leu Pro  
 325 330 335

30 Val Ser Ala Pro Val Pro Ala Ser Val Thr Ala Pro Gly Ser Leu Ser  
 340 345 350

Ala Val Ser Thr Ser Ser Glu Tyr Met Gly Gly Ser Ala Ala Ile Gly  
 355 360 365

35 Pro Ile Thr Pro Ala Thr Thr Ser Ser Ile Thr Ala Ala Val Thr Ala  
 370 375 380

Ser Ser Thr Thr Ser Ala Val Pro Met Gly Asn Gly Val Gly Val Gly  
 385 390 395 400

40 Val Gly Val Gly Gly Asn Val Ser Met Tyr Ala Asn Ala Gln Thr Ala  
 405 410 415

Met Ala Leu Met Gly Val Ala Leu His Ser His Gln Glu Gln Leu Ile  
 420 425 430

45 Gly Gly Val Ala Val Lys Ser Glu His Ser Thr Thr Ala  
 435 440 445

&lt;210&gt; 20

&lt;211&gt; 323

50 &lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;400&gt; 20

55 Arg Pro Glu Cys Val Val Pro Glu Asn Gln Cys Ala Met Lys Arg Arg

1	5	10	15
5	Glu Lys Lys Ala Gln Lys Glu Lys Asp Lys Met Thr Thr Ser Pro Ser 20 25 30		
Ser Gln His Gly Gly Asn Gly Ser Leu Ala Ser Gly Gly Gln Asp 35 40 45			
10	Phe Val Lys Lys Glu Ile Leu Asp Leu Met Thr Cys Glu Pro Pro Gln 50 55 60		
His Ala Thr Ile Pro Leu Leu Pro Asp Glu Ile Leu Ala Lys Cys Gln 65 70 75 80			
15	Ala Arg Asn Ile Pro Ser Leu Thr Tyr Asn Gln Leu Ala Val Ile Tyr 85 90 95		
Lys Leu Ile Trp Tyr Gln Asp Gly Tyr Glu Gln Pro Ser Glu Glu Asp 100 105 110			
Leu Arg Arg Ile Met Ser Gln Pro Asp Glu Asn Glu Ser Gln Thr Asp 115 120 125			
20	Val Ser Phe Arg His Ile Thr Glu Ile Thr Ile Leu Thr Val Gln Leu 130 135 140		
Ile Val Glu Phe Ala Lys Gly Leu Pro Ala Phe Thr Lys Ile Pro Gln 145 150 155 160			
25	Glu Asp Gln Ile Thr Leu Leu Lys Ala Cys Ser Ser Glu Val Met Met 165 170 175		
Leu Arg Met Ala Arg Arg Tyr Asp His Ser Ser Asp Ser Ile Phe Phe 180 185 190			
30	Ala Asn Asn Arg Ser Tyr Thr Arg Asp Ser Tyr Lys Met Ala Gly Met 195 200 205		
Ala Asp Asn Ile Glu Asp Leu Leu His Phe Cys Arg Gln Met Phe Ser 210 215 220			
35	Met Lys Val Asp Asn Val Glu Tyr Ala Leu Leu Thr Ala Ile Val Ile 225 230 235 240		
Phe Ser Asp Arg Pro Gly Leu Glu Lys Ala Gln Leu Val Glu Ala Ile 245 250 255			
40	Gln Ser Tyr Tyr Ile Asp Thr Leu Arg Ile Tyr Ile Leu Asn Arg His 260 265 270		
Cys Gly Asp Ser Met Ser Leu Val Phe Tyr Ala Lys Leu Leu Ser Ile 275 280 285			
45	Leu Thr Glu Leu Arg Thr Leu Gly Asn Gln Asn Ala Glu Met Cys Phe 290 295 300		
Ser Leu Lys Leu Lys Asn Arg Lys Leu Pro Lys Phe Leu Glu Glu Ile 305 310 315 320			
50	Trp Asp Val		

&lt;210&gt; 21

&lt;211&gt; 987

&lt;212&gt; DNA

55 &lt;213&gt; Artificial Sequence

&lt;400&gt; 21

	tgtgctatct	gtggggaccg	c当地ttaggc	aaacactatg	gggtatacag	ttgtgagggc	60
5	tgcaagggtct	tcttcagag	gacagtacgc	aaagacctga	c当地acac	ccgagacaac	120
	aaggactgcc	tgatcgacaa	gagacagcgg	aaaccgtgtc	agta	ctaccagaag	180
	tgccctggcca	tgcccattgaa	gccccggac	gtgcaggagg	agccgcagc	ggcaaggac	240
	cgaaatgaga	acgagggtgaa	gttccaccagc	agtgccaa	aggacatg	cc ttgatagaag	300
10	attcttggaa	cccgagcttc	tgtcgagccc	aagactgaga	catacgttga	ggcaaacatg	360
	gggctgttac	ccagcttacc	aatatggccct	gttaccaaca	tctgtcaagc	agcagacaag	420
	cagctttca	cttcttggaa	gtggccaa	aggatcccac	acttttttgc	gctgccccca	480
15	gacgaccagg	tcatcttgc	acgggcaggc	tggaa	ccggcggc	tgctgtatgc	540
	caccgcgtca	tagctgtgaa	agatggggat	c当地tggcc	ccggccgtca	cgtacaccgg	600
	aacacgcgtc	acagtgttgg	ggtggggcgc	atcttgcaca	gggtgtcaac	agagcttgg	660
20	tctaagatgc	gtgacatgc	gatggacaag	acggagactgg	gctgcctgg	agccatttgc	720
	ctgttcaacc	ctgactctaa	ggggctctca	aaaccctgt	agggtggaggc	tttgaggag	780
	aagggtgtatg	cgtcaactaga	agcgtactgc	aaacacaagt	accctgagca	ggccggcagg	840
25	tttgccttacc	tgctgtctcg	cctgcctgc	ctgcgttcca	tcgggc	tttgccttgg	900
	cacccgttct	tcttcagact	catacggggac	acgcccacatg	acaccttct	catggagatg	960
	ctggaggcac	cacatcaagc	cacccat				987

<210> 22  
 <211> 789  
 <212> DNA  
 <213> Artificial Sequence  
 <400> 22

30	aaggccggaa	ctgtgcagga	ggagccggcag	cggggcaagg	accggaaatga	gaacgagg	60	
	gagtcacca	gcagtgc	ccaa	cgaggacatg	c当地ttagaga	agatttgc	120	
40	gctgtcagc	ccaa	gactga	gacatacgt	gaggcaaa	tggggctgaa	cccccacgtca	180
	ccaaatgacc	ctgttacca	catctgt	caa	gac	ccatcttt	cacttgg	240
	gagtgccca	agaggatccc	acacttttct	gagtcgtccc	tagacgacca	ggtcata	cctg	300
45	ctacgggcag	gctggaa	cg	gtgtgtatc	gcctcc	cccacccgtc	cata	360
	aaagatggaa	ttctcttggc	caccggc	ctgtacacc	ggaac	acgcgc	tcac	420
	gggggtggc	ccatcttgc	cagggtgt	ctaa	acagac	gttgc	gat	480
50	catggaca	agacggact	gggtgtctg	cgacccat	ttctgtca	ccctgactt		540
	aagggtct	caaaccctgc	tgagggtgg	gcgttggag	agaagg	gtgtca	acta	600
	gaacgtact	gcaaa	acacaa	gtaccctg	cagccggc	ggtttgc	ca	660

	ccctgcctg cactgcgttc catcggtc aagtgcctgg agcacctgtt cttcttcaag	720
	ctcatcgaaa acacgccccat cgacacccat ctcataggaga tgctggaggc accacatcaa	780
5	gccacccatag	789

<210> 23  
<211> 714  
10 <212> DNA  
<213> Artificial Sequence  
<400> 23

15	gccaacgagg acatgcgtt agagaagatt ctggaaagccg agcttgcgtt cgagccaaag	60
	actgagacat acgtggaggc aaacatgggg ctgaacccca gtcacccaa tgaccctgtt	120
	. accaacatct gtcaagcgc agacaaggcag ctcttcactc ttgtggagtggcacaagagg	180
20	atcccacact ttctcttagt gccccttagac gaccaggcgtt ctctgcgtt ggaggctgg	240
	aacgagctgc tgatgcctc ctctccac cgttccatag ctgtaaaga tgggatttctc	300
	ctggccacccg gcctgcacgt acacccgaac agcgctcaca gtgctgggtt gggcgccatc	360
25	tttgacaggg tgcttaacaga gctgggttct aagatgcgtt acatgcatggacaagacg	420
	gagctgggtt gcctgcgacg cattgtcctg ttcaacccctg actctaagg gctctcaaac	480
	cctgctgagg tggaggcgtt gagggagaag gtgtatgcgt cactagaacg gtactgcaaa	540
30	cacaagtacc ctgagcagcc gggcagggtt gccaagctgc tgctccgcgtt gctgcactg	600
	cgttccatcg ggctcaatgtt cctggaggcac ctgttttct tcaagtcatacgccggacac	660
	cccatcgaca ctttcctcat ggagatgtt gaggcaccac atcaaggccac ctatgtt	714

<210> 24  
<211> 536  
15 <212> DNA  
<213> Artificial Sequence  
<400> 24

35	ggatccccaca cttttctgat ctggcccttag acgaccaggat ctcctgcata cgggcaggct	60
	ggaaacgaggat gctgtatgcctc tccttctccc accgcgtccat agctgtgaaa gatggggatc	120
40	tcctggccac cggcgtgcac gtacaccggaa acacgcgtca cagtgtgggg gtggggccca	180
	tctttgacat ggtgttcaaca gagctgggtt ctaagatgcgt tgacatgcgtt atggacaaga	240
	cgggactggg ctgcctgcgtt gccaatgttcc ttgttcaaccc ttgttcaatgggtctcaaa	300
45	acccctgtgtt ggtggaggcgtt ttggaggaga aggtgtatgcgtt cactagaacgcgtactgtca	360
	aacacaagttt ccctggaggcgtt cggggcagggtt tgcccaatgtt gctgtccgcgtt ctcgttgcac	420
	tgctgttccatcg cgggttcaagtt tgccgttgcgtt acctgtttttt cttcaatgtt atcggttgcac	480
50	cgcccatcgaca caccccttcatacgatgtt cttggaggcaccac atcatcaaggccac acctatgtt	536

<210> 25  
<211> 672

<212> DNA  
<213> Artificial Sequence  
<400> 25

5	gccaacgagg acatgcctgt agagaagatt ctggaaagccg agcttgcgtg cgagcccaag	60
	actgagacat acgtggaggc aaacatgggg ctgaacccca gctcacccaa tgaccctgtt	120
	accaacatct gtcaagcgcg agacaagcg ctcttcactc ttgtggagtggccaaagg	180
10	atccccacact ttctcgatg gccccttagac gaccagggtca tcctgtcttcgg gcaggctgg	240
	aacgagctgc tgatcgctc cttctccac cgctccatag ctgtgaaaga tggattctc	300
	ctggccacccg gcctgcacgt acaccggaaac agcgctcaca gtgtctgggtt gggcgcacatc	360
15	tttgacagggt tgctaacaga gctgggtct aagatgcgtg acatgcagat ggacaagacg	420
	gagctggct gcctgcgagc cattgtctcg ttcaaccctg actctaaggg gctctcaaacc	480
	cctgtgtgagg tggaggcggtt gagggagaag gtgtatgcgt cactagaagc gtactgc当地	540
20	cacaagtacc ctgagcagcc gggcagggtt gccaagctgc tgctccgcct gcctgcactg	600
	cgttccatcg ggctcaagtgc cttggaggcac ctgttcttc tcaagctcat cggggcacacg	660
	cccatcgaca cc	672

25 <210> 26  
<211> 1123  
<212> DNA  
<213> Artificial Sequence  
30 <400> 26

	tgcggccatct	gccccggaccc	ctcctcaggc	aagcaactatg	gagtgtacag	ctgcgcgggg	60
	tgcaagggt	tcttcggcg	gacgggtcgcc	aaggacctga	cctacacctg	ccgcgcacaac	120
35	aaggactgcc	tgattgcaca	ccggcagccg	aaccgggtcc	agtaactgccc	cttaccagaag	180
	tgcctggcca	tggggcatgaa	ccgggaaagcc	gtgcaggagg	agccggcagcg	tggcaaggac	240
	cgaaacgaga	atggagggtgg	gtgcgaccaggc	agcgcaccaacg	aggacatgcc	ggttggagagg	300
40	atcctggagg	ctgagctggc	ctggggagccc	aagaccgaga	cctacgtgg	ggccaaacatg	360
	gggctgaacc	ccagctcgcc	qaacgacccct	gtcacaaca	tttgcacaa	agccgcacaaa	420
	cagctttca	ccctgggtgaa	gtggggcaag	cggtatccac	actttcteaga	gtgcgccttg	480
45	gacgaccagg	tcatctctgc	ccggggcaggc	tggatgtggc	tgctcatcgcc	ctctttctcc	540
	caccgcctca	tcggcgtgaa	ggacgggatc	ctctggcca	ccgggctgca	cgttccacccg	600
	aacagcgc	ccacagcgcagg	ggtgtggcc	atctttgaca	gggtgtctgac	ggagcttgtg	660
50	tccaaagatgc	gggacatgca	gttggacaag	acggagctgg	gtctgcctcg	cgccatcgtc	720
	ctcttttaacc	ctgactccaa	ggggctctcg	aaccggcc	agttggaggc	gtctggaggag	780
	aaggctatg	cgtccttgg	ggctctactgc	aagcacaagt	acccagagca	ggccggaaagg	840
55	ttcgctaa	gtttgtccg	cttcggccgt	ctcgctcc	tcgggctcaa	atgccttggaa	900

	catctttct tcttcaagct catcggggac acaccattg acaccttct tatggatgt	960
5	ctggaggcgc cgacacaaat gacttagccc tgccggccca tcctttgtgc ccaccgttc	1020
	tggccacccct gectggacgc cagctgttct tctcagcctg agccctgtcc ctgccttct	1080
	ctgcctggcc ttttttgact ttggggcaca gcctgtcaact gct	1123
10	<210> 27	
	<211> 925	
	<212> DNA	
	<213> Artificial Sequence	
	<400> 27	
15	aagcgggaag ccgtgcagga ggagcggcag cgtggcaagg accggaacga gaatgaggta	60
	gagtgcacca gcagcgc当地 cgaggacatg ccgggtggaga ggatcttggg ggctgagctg	120
20	gccgtggcgc ccaagacccg gacctacgtt gaggccaaaca tggggctgaa ccccaagctg	180
	ccgaacgacc ctgtcaccaa catttgc当地 gcagccgaca aacagttttt cacccttgg	240
	gagtggccca agcggttccc acacttctca gagctgtcccc tggacgacca ggtcatctg	300
25	ctgcgggcag gcttggatgt gctgttccatc gcctttctt cccaccgc当地 catgcgc当地	360
	aaggacggga ttcttcttgc caccgggtgtt cacgttccacc ggaacgc当地 ccacagc当地	420
	gggggtggccg ccatttttgc cagggtgtgtt acgggatgtt tgcttccatg gggggacatg	480
30	cagatggacca agacggagctt gggctgttcc cgcgc当地 catcgttccatg ttcttccatg	540
	aaggggctctt cgaacccggc cggagggtggag ggcgttgggg agaaggctta tgctgttccatg	600
	gaggccatctt gcaagcacaat gtacccttgc cagccggaa ggttc当地 gtc当地 ttgttgc当地	660
35	cgcccttccgg ctcttgc当地 catccggc当地 aaatggcttcc aacatcttctt ttcttccatg	720
	ctcatcgccggg acacacccat tgacacccttcc ttatggaga tgctggaggc gccgc当地	780
	atgacttagg cctcgccggc catctttgtt gcccacccgtt tcttgc当地 acccttgc当地	840
40	gcacagctgtt ttcttc当地 ctttgc当地 tgaggccctgtt cccttgc当地 ttcttgc当地	900
	cttggggccca cagcctgtca ctgtt	925
45	<210> 28	
	<211> 850	
	<212> DNA	
	<213> Artificial Sequence	
	<400> 28	
50	gcacacggagg acatggccgtt ggagaggatc ctggaggctg agctggccgtt ggagcccaag	60
	accggagactt acgtggaggc当地 aaacatgggg ctgaacccttcc gctcgccgaa ccaccctgtc	120
	accacacattt gccaaggc当地 cgacaaaacag cttttccatcc ttgtggagggtt ggccaaaggc当地	180
55	atccccacact tctcagagctt gccc当地 ttcttgc当地 gaccaggatc ttcttgc当地	240
	aatgagctgc tcatcgccctt ctttccatcc cgttccatg ccgtgaaggaa cgggatcc	300

	ctggccacccg ggctgcacgt ccaccgaaac agcgcaccaca ggcgcagggtt gggcgccatc	360
5	tttgacaggg tgctgacgga gcttgttcc aagatgcggg acatgcagat ggacaagacg	420
	gagctgggt gcctgcgcgc categtcctc tttaaccctg actccaagg gctctcgaaac	480
	ccggccgagg tggaggcgct gagggagaag gtctatgcgt ctgtggaggc ctactgcaag	540
	cacaagtacc cagagcagcc gggaaagggtc gctaagctct tgctccgcgt gccggctctg	600
10	cgctccatcg ggctcaaatacg cctggaaacat ctcttcttct tcaagctcat cggggacaca	660
	ccattgaca ccttccttat ggagatgcgt gaggcgccgc accaaatgac ttaggcctgc	720
	ggcccatcc tttgtccccca ccggttctgg ccaccctgccc tggacgccag ctgttcttct	780
15	cagctgagc cctgtccctg cccttctctg cctggccctgt ttggactttg gggcacagcc	840
	tgtcaactgct	850
	<210> 29	
20	<211> 670	
	<212> DNA	
	<213> Artificial Sequence	
	<400> 29	
	atccccacact ttcagagct gcccctggac gaccagggtca tcctgtgcgt ggcaggctgg	60
25	aatggagctgc tcategcctc ctctccac cgctccatcg ccgtgaagga cgggatcc	120
	ctggccacccg ggctgcacgt ccaccgaaac agcgcaccaca ggcgcagggtt gggcgccatc	180
30	tttgacaggg tgctgacgga gcttgttcc aagatgcggg acatgcagat ggacaagacg	240
	gagctgggt gcctgcgcgc catgccttc tttaaccctg actccaagg gctctcgaaac	300
	ccggccgagg tggaggcgct gagggagaag gtctatgcgt ctgtggaggc ctactgcaag	360
35	cacaagtacc cagagcagcc gggaaagggtc gctaagctct tgctccgcgt gccggctctg	420
	cgctccatcg ggctcaaatacg cctggaaacat ctcttcttct tcaagctcat cggggacaca	480
	ccattgaca ccttccttat ggagatgcgt gaggcgccgc accaaatgac ttaggcctgc	540
40	ggcccatcc tttgtccccca ccggttctgg ccaccctgccc tggacgccag ctgttcttct	600
	cagctgagc cctgtccctg cccttctctg cctggccctgt ttggactttg gggcacagcc	660
	tgtcaactgct	670
45	<210> 30	
	<211> 672	
	<212> DNA	
	<213> Artificial Sequence	
50	<400> 30	
	gccaacgagg acatgcgggt ggagaggatc ctggaggctg agctggccgtt ggagcccaag	60
55	accgagacct acgtggggc aaacatgggg ctgaacccca gtcgcggaa cgaccctgtc	120
	accaacattt gccaagcage cgacaaaacag ctttacccc tggtgagtg ggcaagcgg	180
	atccccacact ttcagagct gcccctggac gaccagggtca tcctgtgcgt ggcaggctgg	240

	aatgagctgc tcatcgccctc cttctccac cgctccatcg ccgtgaaggga cgggatccctc	300
	ctggccacccg ggctgcacgt ccacccggaa acgcgccaca ggcgcgggggt gggcgccatc	360
5	tttgacaggg tgctgacgga gcttgggttc aagatgcggg acatgcagat ggacaagacg	420
	gagctgggt gcctgcgcgc catcgccctc ttaaccctg actccaaggg gtctcgac	480
	ccggccgagg tggaggcgct gagggagaag gtctatgcgt ccttgaggc ctactgcaag	540
10	cacaagtacc cagagcagcc gggaaaggttc gctaagctct tgctccgcct gccggctctg	600
	cgctccatcg ggctcaaatacg cctgaaacat ctcttttct tcaagtcat cggggacaca	660
	cccatggaca cc	672

15

&lt;210&gt; 31

&lt;211&gt; 328

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

20

&lt;400&gt; 31

Cys	Ala	Ile	Cys	Gly	Asp	Arg	Ser	Ser	Gly	Lys	His	Tyr	Gly	Val	Tyr
1				5					10				15		
Ser	Cys	Glu	Gly	Cys	Lys	Gly	Phe	Phe	Lys	Arg	Thr	Val	Arg	Lys	Asp
		20						25					30		
Leu	Thr	Tyr	Thr	Cys	Arg	Asp	Asn	Lys	Asp	Cys	Leu	Ile	Asp	Lys	Arg
				35				40				45			
Gln	Arg	Asn	Arg	Cys	Gln	Tyr	Cys	Arg	Tyr	Gln	Lys	Cys	Leu	Ala	Met
				50			55				60				
Gly	Met	Lys	Arg	Glu	Ala	Val	Gln	Glu	Glu	Arg	Gln	Arg	Gly	Lys	Asp
				65			70				75				80
Arg	Asn	Glu	Asn	Glu	Val	Glu	Ser	Thr	Ser	Ser	Ala	Asn	Glu	Asp	Met
					85				90				95		
Pro	Val	Glu	Lys	Ile	Leu	Glu	Ala	Glu	Leu	Ala	Val	Glu	Pro	Lys	Thr
				100					105				110		
Glu	Thr	Tyr	Val	Glu	Ala	Asn	Met	Gly	Leu	Asn	Pro	Ser	Ser	Pro	Asn
				115			120				125				
Asp	Pro	Val	Thr	Asn	Ile	Cys	Gln	Ala	Ala	Asp	Lys	Gln	Leu	Phe	Thr
				130			135				140				
Leu	Val	Glu	Trp	Ala	Lys	Arg	Ile	Pro	His	Phe	Ser	Glu	Leu	Pro	Leu
				145			150				155				160
Asp	Asp	Gln	Val	Ile	Leu	Leu	Arg	Ala	Gly	Trp	Asn	Glu	Leu	Leu	Ile
				165				170					175		
Ala	Ser	Phe	Ser	His	Arg	Ser	Ile	Ala	Val	Lys	Asp	Gly	Ile	Leu	Leu
				180				185					190		
Ala	Thr	Gly	Leu	His	Val	His	Arg	Asn	Ser	Ala	His	Ser	Ala	Gly	Val
				195				200				205			
Gly	Ala	Ile	Phe	Asp	Arg	Val	Leu	Thr	Glu	Leu	Val	Ser	Lys	Met	Arg

55

210

215

220

Asp Met Gln Met Asp Lys Thr Glu Leu Gly Cys Leu Arg Ala Ile Val  
 225 230 235 240

5 Leu Phe Asn Pro Asp Ser Lys Gly Leu Ser Asn Pro Ala Glu Val Glu  
 245 250 255

Ala Leu Arg Glu Lys Val Tyr Ala Ser Leu Glu Ala Tyr Cys Lys His  
 260 265 270

10 Lys Tyr Pro Glu Gln Pro Gly Arg Phe Ala Lys Leu Leu Arg Leu  
 275 280 285

Pro Ala Leu Arg Ser Ile Gly Leu Lys Cys Leu Glu His Leu Phe Phe  
 290 295 300

15 Phe Lys Leu Ile Gly Asp Thr Pro Ile Asp Thr Phe Leu Met Glu Met  
 305 310 315 320

Leu Glu Ala Pro His Gln Ala Thr  
 325

20 <210> 32  
 <211> 262  
 <212> PRT  
 <213> Artificial Sequence  
 25 <400> 32

Lys Arg Glu Ala Val Gln Glu Glu Arg Gln Arg Gly Lys Asp Arg Asn  
 1 5 10 15

30 Glu Asn Glu Val Glu Ser Thr Ser Ser Ala Asn Glu Asp Met Pro Val  
 20 25 30

Glu Lys Ile Leu Glu Ala Glu Leu Ala Val Glu Pro Lys Thr Glu Thr  
 35 40 45

35 Tyr Val Glu Ala Asn Met Gly Leu Asn Pro Ser Ser Pro Asn Asp Pro  
 50 55 60

Val Thr Asn Ile Cys Gln Ala Ala Asp Lys Gln Leu Phe Thr Leu Val  
 65 70 75 80

40 Glu Trp Ala Lys Arg Ile Pro His Phe Ser Glu Leu Pro Leu Asp Asp  
 85 90 95

Gln Val Ile Leu Leu Arg Ala Gly Trp Asn Glu Leu Leu Ile Ala Ser  
 100 105 110

45 Phe Ser His Arg Ser Ile Ala Val Lys Asp Gly Ile Leu Leu Ala Thr  
 115 120 125

Gly Leu His Val His Arg Asn Ser Ala His Ser Ala Gly Val Gly Ala  
 130 135 140

50 Ile Phe Asp Arg Val Leu Thr Glu Leu Val Ser Lys Met Arg Asp Met  
 145 150 155 160

Gln Met Asp Lys Thr Glu Leu Gly Cys Leu Arg Ala Ile Val Leu Phe  
 165 170 175

55 Asn Pro Asp Ser Lys Gly Leu Ser Asn Pro Ala Glu Val Glu Ala Leu  
 180 185 190

Arg Glu Lys Val Tyr Ala Ser Leu Glu Ala Tyr Cys Lys His Lys Tyr  
 195 200 205  
 Pro Glu Gln Pro Gly Arg Phe Ala Lys Leu Leu Leu Arg Leu Pro Ala  
 210 215 220  
 Leu Arg Ser Ile Gly Leu Lys Cys Leu Glu His Leu Phe Phe Phe Lys  
 225 230 235 240  
 Leu Ile Gly Asp Thr Pro Ile Asp Thr Phe Leu Met Glu Met Leu Glu  
 245 250 255  
 Ala Pro His Gln Ala Thr  
 260

15 <210> 33  
 <211> 237  
 <212> PRT  
 <213> Artificial Sequence  
 <400> 33

20 Ala Asn Glu Asp Met Pro Val Glu Lys Ile Leu Glu Ala Glu Leu Ala  
 1 5 10 15  
 Val Glu Pro Lys Thr Glu Thr Tyr Val Glu Ala Asn Met Gly Leu Asn  
 25 20 25 30  
 Pro Ser Ser Pro Asn Asp Pro Val Thr Asn Ile Cys Gln Ala Ala Asp  
 35 40 45  
 Lys Gln Leu Phe Thr Leu Val Glu Trp Ala Lys Arg Ile Pro His Phe  
 30 50 55 60  
 Ser Glu Leu Pro Leu Asp Asp Gln Val Ile Leu Leu Arg Ala Gly Trp  
 65 70 75 80  
 Asn Glu Leu Leu Ile Ala Ser Phe Ser His Arg Ser Ile Ala Val Lys  
 85 90 95  
 35 Asp Gly Ile Leu Leu Ala Thr Gly Leu His Val His Arg Asn Ser Ala  
 100 105 110  
 His Ser Ala Gly Val Gly Ala Ile Phe Asp Arg Val Leu Thr Glu Leu  
 40 115 120 125  
 Val Ser Lys Met Arg Asp Met Gln Met Asp Lys Thr Glu Leu Gly Cys  
 130 135 140  
 Leu Arg Ala Ile Val Leu Phe Asn Pro Asp Ser Lys Gly Leu Ser Asn  
 145 150 155 160  
 45 Pro Ala Glu Val Glu Ala Leu Arg Glu Lys Val Tyr Ala Ser Leu Glu  
 165 170 175  
 Ala Tyr Cys Lys His Lys Tyr Pro Glu Gln Pro Gly Arg Phe Ala Lys  
 180 185 190  
 50 Leu Leu Leu Arg Leu Pro Ala Leu Arg Ser Ile Gly Leu Lys Cys Leu  
 195 200 205  
 Glu His Leu Phe Phe Lys Leu Ile Gly Asp Thr Pro Ile Asp Thr  
 210 215 220

Phe Leu Met Glu Met Leu Glu Ala Pro His Gln Ala Thr  
 225   230   235

5   <210> 34  
 <211> 177  
 <212> PRT  
 <213> Artificial Sequence  
 <400> 34

10

Ile Pro His Phe Ser Glu Leu Pro Leu Asp Asp Gln Val Ile Leu Leu  
 1   5   10   15  
 Arg Ala Gly Trp Asn Glu Leu Leu Ile Ala Ser Phe Ser His Arg Ser  
 15   20   25   30  
 Ile Ala Val Lys Asp Gly Ile Leu Leu Ala Thr Gly Leu His Val His  
 20   35   40   45  
 Arg Asn Ser Ala His Ser Ala Gly Val Gly Ala Ile Phe Asp Arg Val  
 25   50   55   60  
 Leu Thr Glu Leu Val Ser Lys Met Arg Asp Met Gln Met Asp Lys Thr  
 30   65   70   75   80  
 Glu Leu Gly Cys Leu Arg Ala Ile Val Leu Phe Asn Pro Asp Ser Lys  
 35   85   90   95  
 Gly Leu Ser Asn Pro Ala Glu Val Glu Ala Leu Arg Glu Lys Val Tyr  
 40   100   105   110  
 Ala Ser Leu Glu Ala Tyr Cys Lys His Lys Tyr Pro Glu Gln Pro Gly  
 45   115   120   125  
 Arg Phe Ala Lys Leu Leu Arg Leu Pro Ala Leu Arg Ser Ile Gly  
 50   130   135   140  
 Leu Lys Cys Leu Glu His Leu Phe Phe Lys Leu Ile Gly Asp Thr  
 55   145   150   155   160  
 Pro Ile Asp Thr Phe Leu Met Glu Met Leu Glu Ala Pro His Gln Ala  
 60   165   170   175

Thr

40   <210> 35  
 <211> 224  
 <212> PRT  
 <213> Artificial Sequence  
 45   <400> 35

Ala Asn Glu Asp Met Pro Val Glu Lys Ile Leu Glu Ala Glu Leu Ala  
 1   5   10   15  
 Val Glu Pro Lys Thr Glu Thr Tyr Val Glu Ala Asn Met Gly Leu Asn  
 50   20   25   30  
 Pro Ser Ser Pro Asn Asp Pro Val Thr Asn Ile Cys Gln Ala Ala Asp  
 55   35   40   45  
 Lys Glu Leu Phe Thr Leu Val Glu Trp Ala Lys Arg Ile Pro His Phe  
 60   50   55   60

Ser Glu Leu Pro Leu Asp Asp Gln Val Ile Leu Leu Arg Ala Gly Trp  
 65 70 75 80

5 Asn Glu Leu Leu Ile Ala Ser Phe Ser His Arg Ser Ile Ala Val Lys  
 85 90 95

Asp Gly Ile Leu Leu Ala Thr Gly Leu His Val His Arg Asn Ser Ala  
 100 105 110

10 His Ser Ala Gly Val Gly Ala Ile Phe Asp Arg Val Leu Thr Glu Leu  
 115 120 125

15 Val Ser Lys Met Arg Asp Met Gln Met Asp Lys Thr Glu Leu Gly Cys  
 130 135 140

Leu Arg Ala Ile Val Leu Phe Asn Pro Asp Ser Lys Gly Leu Ser Asn  
 145 150 155 160

Pro Ala Glu Val Glu Ala Leu Arg Glu Lys Val Tyr Ala Ser Leu Glu  
 165 170 175

20 Ala Tyr Cys Lys His Lys Tyr Pro Glu Gln Pro Gly Arg Phe Ala Lys  
 180 185 190

Leu Leu Leu Arg Leu Pro Ala Leu Arg Ser Ile Gly Leu Lys Cys Leu  
 195 200 205

25 Glu His Leu Phe Phe Lys Leu Ile Gly Asp Thr Pro Ile Asp Thr  
 210 215 220

&lt;210&gt; 36

&lt;211&gt; 328

&lt;212&gt; PRT

30 &lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;221&gt; misc feature

&lt;223&gt; Novel Sequence

&lt;400&gt; 36

35

Cys Ala Ile Cys Gly Asp Arg Ser Ser Gly Lys His Tyr Gly Val Tyr  
 1 5 10 15

40

Ser Cys Glu Gly Cys Lys Gly Phe Phe Lys Arg Thr Val Arg Lys Asp  
 20 25 30

45

Leu Thr Tyr Thr Cys Arg Asp Asn Lys Asp Cys Leu Ile Asp Lys Arg  
 35 40 45

50

Gln Arg Asn Arg Cys Gln Tyr Cys Arg Tyr Gln Lys Cys Leu Ala Met  
 50 55 60

55

Gly Met Lys Arg Glu Ala Val Gln Glu Glu Arg Gln Arg Gly Lys Asp  
 65 70 75 80

Arg Asn Glu Asn Glu Val Glu Ser Thr Ser Ser Ala Asn Glu Asp Met  
 85 90 95

Pro Val Glu Arg Ile Leu Glu Ala Glu Leu Ala Val Glu Pro Lys Thr  
 100 105 110

Glu Thr Tyr Val Glu Ala Asn Met Gly Leu Asn Pro Ser Ser Pro Asn  
 115 120 125

Asp Pro Val Thr Asn Ile Cys Gln Ala Ala Asp Lys Gln Leu Phe Thr  
 130 135 140  
 Leu Val Glu Trp Ala Lys Arg Ile Pro His Phe Ser Glu Leu Pro Leu  
 145 150 155 160  
 Asp Asp Gln Val Ile Leu Leu Arg Ala Gly Trp Asn Glu Leu Leu Ile  
 165 170 175  
 Ala Ser Phe Ser His Arg Ser Ile Ala Val Lys Asp Gly Ile Leu Leu  
 180 185 190  
 Ala Thr Gly Leu His Val His Arg Asn Ser Ala His Ser Ala Gly Val  
 195 200 205  
 Gly Ala Ile Phe Asp Arg Val Leu Thr Glu Leu Val Ser Lys Met Arg  
 210 215 220  
 Asp Met Gln Met Asp Lys Thr Glu Leu Gly Cys Leu Arg Ala Ile Val  
 225 230 235 240  
 Leu Phe Asn Pro Asp Ser Lys Gly Leu Ser Asn Pro Ala Glu Val Leu  
 245 250 255  
 Ala Leu Arg Glu Lys Val Tyr Ala Ser Leu Glu Ala Tyr Cys Lys His  
 260 265 270  
 Lys Tyr Pro Glu Gln Pro Gly Arg Phe Ala Lys Leu Leu Arg Leu  
 275 280 285  
 Pro Ala Leu Arg Ser Ile Gly Leu Lys Cys Leu Glu His Leu Phe Phe  
 290 295 300  
 Phe Lys Leu Ile Gly Asp Thr Pro Ile Asp Thr Phe Leu Met Glu Met  
 305 310 315 320  
 Leu Glu Ala Pro His Gln Met Thr  
 325

<210> 37  
 <211> 262  
 <212> PRT  
 <213> Artificial Sequence  
 <220>  
 <221> misc feature  
 <223> Novel Sequence  
 <400> 37

Lys Arg Glu Ala Val Gln Glu Glu Arg Gln Arg Gly Lys Asp Arg Asn  
 1 5 10 15  
 Glu Asn Glu Val Glu Ser Thr Ser Ser Ala Asn Glu Asp Met Pro Val  
 20 25 30  
 Glu Arg Ile Leu Glu Ala Glu Leu Ala Val Glu Pro Lys Thr Glu Thr  
 35 40 45  
 Tyr Val Glu Ala Asn Met Gly Leu Asn Pro Ser Ser Pro Asn Asp Pro  
 50 55 60  
 Val Thr Asn Ile Cys Gln Ala Ala Asp Lys Gln Leu Phe Thr Leu Val

65	70	. 75		
	Glu Trp Ala Lys Arg Ile Pro His Phe Ser Glu Leu Pro Leu Asp Asp			
	85	90	95	
5	Gln Val Ile Leu Leu Arg Ala Gly Trp Asn Glu Leu Leu Ile Ala Ser			
	100	105	110	
	Phe Ser His Arg Ser Ile Ala Val Lys Asp Gly Ile Leu Leu Ala Thr			
	115	120	125	
10	Gly Leu His Val His Arg Asn Ser Ala His Ser Ala Gly Val Gly Ala			
	130	135	140	
	Ile Phe Asp Arg Val Leu Thr Glu Leu Val Ser Lys Met Arg Asp Met			
	145	150	155	160
15	Gln Met Asp Lys Thr Glu Leu Gly Cys Leu Arg Ala Ile Val Leu Phe			
	165	170	175	
	Asn Pro Asp Ser Lys Gly Leu Ser Asn Pro Ala Glu Val Glu Ala Leu			
	180	185	190	
20	Arg Glu Lys Val Tyr Ala Ser Leu Glu Ala Tyr Cys Lys His Lys Tyr			
	195	200	205	
	Pro Glu Gln Pro Gly Arg Phe Ala Lys Leu Leu Leu Arg Leu Pro Ala			
	210	215	220	
25	Leu Arg Ser Ile Gly Leu Lys Cys Leu Glu His Leu Phe Phe Phe Lys			
	225	230	235	240
	Leu Ile Gly Asp Thr Pro Ile Asp Thr Phe Leu Met Glu Met Leu Glu			
	245	250	255	
30	Ala Pro His Gln Met Thr			
	260			

&lt;210&gt; 38

&lt;211&gt; 237

35 &lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;221&gt; misc feature

&lt;223&gt; Novel Sequence

40 &lt;400&gt; 38

45	Ala Asn Glu Asp Met Pro Val Glu Arg Ile Leu Glu Ala Glu Leu Ala			
	1	5	10	15
	Val Glu Pro Lys Thr Glu Thr Tyr Val Glu Ala Asn Met Gly Leu Asn			
	20	25	30	
	Pro Ser Ser Pro Asn Asp Pro Val Thr Asn Ile Cys Gln Ala Ala Asp			
	35	40	45	
50	Lys Gln Leu Phe Thr Leu Val Glu Trp Ala Lys Arg Ile Pro His Phe			
	50	55	60	
	Ser Glu Leu Pro Leu Asp Asp Gln Val Ile Leu Leu Arg Ala Gly Trp			
	65	70	75	80

Asn Glu Leu Leu Ile Ala Ser Phe Ser His Arg Ser Ile Ala Val Lys  
 85 90 95

Asp Gly Ile Leu Leu Ala Thr Gly Leu His Val His Arg Asn Ser Ala  
 100 105 110

His Ser Ala Gly Val Gly Ala Ile Phe Asp Arg Val Leu Thr Glu Leu  
 115 120 125

Val Ser Lys Met Arg Asp Met Gln Met Asp Lys Thr Glu Leu Gly Cys  
 130 135 140

Leu Arg Ala Ile Val Leu Phe Asn Pro Asp Ser Lys Gly Leu Ser Asn  
 145 150 155 160

Pro Ala Glu Val Glu Ala Leu Arg Glu Lys Val Tyr Ala Ser Leu Glu  
 165 170 175

Ala Tyr Cys Lys His Lys Tyr Pro Glu Gln Pro Gly Arg Phe Ala Lys  
 180 185 190

Leu Leu Leu Arg Leu Pro Ala Leu Arg Ser Ile Gly Leu Lys Cys Leu  
 195 200

Glu His Leu Phe Phe Lys Leu Ile Gly Asp Thr Pro Ile Asp Thr  
 210 215 220

Phe Leu Met Glu Met Leu Glu Ala Pro His Gln Met Thr  
 225 230 235

25

&lt;210&gt; 39

&lt;211&gt; 177

&lt;212&gt; PRT

30

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;221&gt; misc\_feature

&lt;223&gt; Novel Sequence

&lt;400&gt; 39

35

Ile Pro His Phe Ser Glu Leu Pro Leu Asp Asp Gln Val Ile Leu Leu  
 1 5 10 15

Arg Ala Gly Trp Asn Glu Leu Leu Ile Ala Ser Phe Ser His Arg Ser  
 20 25 30

Ile Ala Val Lys Asp Gly Ile Leu Leu Ala Thr Gly Leu His Val His  
 35 40 45

Arg Asn Ser Ala His Ser Ala Gly Val Gly Ala Ile Phe Asp Arg Val  
 50 55 60

Leu Thr Glu Leu Val Ser Lys Met Arg Asp Met Gln Met Asp Lys Thr  
 65 70 75 80

Glu Leu Gly Cys Leu Arg Ala Ile Val Leu Phe Asn Pro Asp Ser Lys  
 85 90 95

Gly Leu Ser Asn Pro Ala Glu Val Glu Ala Leu Arg Glu Lys Val Tyr  
 100 105 110

Ala Ser Leu Glu Ala Tyr Cys Lys His Lys Tyr Pro Glu Gln Pro Gly  
 115 120 125

55

Arg Phe Ala Lys Leu Leu Leu Arg Leu Pro Ala Leu Arg Ser Ile Gly  
 130 135 140  
 5 Leu Lys Cys Leu Glu His Leu Phe Phe Phe Lys Leu Ile Gly Asp Thr  
 145 150 155 160  
 Pro Ile Asp Thr Phe Leu Met Glu Met Leu Glu Ala Pro His Gln Met  
 165 170 175

## Thr

10

<210> 40  
 <211> 224  
 <212> PRT  
 15 <213> Artificial Sequence  
 <220>  
 <221> misc feature  
 <223> Novel Sequence  
 <400> 40

20

Ala Asn Glu Asp Met Pro Val Glu Arg Ile Leu Glu Ala Glu Leu Ala  
 1 5 10 15  
 25 Val Glu Pro Lys Thr Glu Thr Tyr Val Glu Ala Asn Met Gly Leu Asn  
 20 25 30  
 Pro Ser Ser Pro Asn Asp Pro Val Thr Asn Ile Cys Gln Ala Ala Asp  
 35 40 45  
 30 Lys Gln Leu Phe Thr Leu Val Glu Trp Ala Lys Arg Ile Pro His Phe  
 50 55 60  
 Ser Glu Leu Pro Leu Asp Asp Gln Val Ile Leu Leu Arg Ala Gly Trp  
 65 70 75 80  
 Asn Glu Leu Leu Ile Ala Ser Phe Ser His Arg Ser Ile Ala Val Lys  
 85 90 95  
 35 Asp Gly Ile Leu Leu Ala Thr Gly Leu His Val His Arg Asn Ser Ala  
 100 105 110  
 His Ser Ala Gly Val Gly Ala Ile Phe Asp Arg Val Leu Thr Glu Leu  
 115 120 125  
 40 Val Ser Lys Met Arg Asp Met Gln Met Asp Lys Thr Glu Leu Gly Cys  
 130 135 140  
 Leu Arg Ala Ile Val Leu Phe Asn Pro Asp Ser Lys Gly Leu Ser Asn  
 145 150 155 160  
 45 Pro Ala Glu Val Glu Ala Leu Arg Glu Lys Val Tyr Ala Ser Leu Glu  
 165 170 175  
 Ala Tyr Cys Lys His Lys Tyr Pro Glu Gln Pro Gly Arg Phe Ala Lys  
 180 185 190  
 50 Leu Leu Leu Arg Leu Pro Ala Leu Arg Ser Ile Gly Leu Lys Cys Leu  
 195 200 205  
 Glu His Leu Phe Phe Phe Lys Leu Ile Gly Asp Thr Pro Ile Asp Thr

55

210

215

220

<210> 41  
 <211> 441  
 <212> DNA  
 <213> Artificial Sequence  
 5 <220>  
 <221> misc feature  
 <223> Novel Sequence  
 <400> 41

10	atgaagctac tgcgttctat cgaacaagca tgcgatattt gccgacttaa aaagctcaag	60
	tgcgtccaaag aaaaaccgaa gtgcgcggaa tgcgtcagaaga acaactgggg gtgtcgctac	120
	tctccaaaaa ccaaaagggtc tccgcgtact agggcacatc tgacagaagt ggaatcaagg	180
15	ctagaagac tggAACAGCT attttctactg atttttcttc gagaAGACCT tgacatgatt	240
	ttggAAATGGG attctttaca ggatataaa gcattgttaa caggattatt tgcataagat	300
	aatgtgaata aagatGCCGT cacagataga ttggcttcgg tggagactga tatgcctcta	360
20	acattgagac agcatagaat aagtgcgaca tcatcatcgg aagagagtag taacaaagg	420
	caaagacagt tgactgtatc g	441

25 <210> 42  
 <211> 147  
 <212> PRT  
 <213> Artificial Sequence  
 <220>  
 30 <221> misc\_feature  
 <223> Novel Sequence  
 <400> 42

35	Met Lys Leu Leu Ser Ser Ile Glu Gln Ala Cys Asp Ile Cys Arg Leu	
	1 5 10 15	
	Lys Lys Leu Lys Cys Ser Lys Glu Lys Pro Lys Cys Ala Lys Cys Leu	
	20 25 30	
40	Lys Asn Asn Trp Glu Cys Arg Tyr Ser Pro Lys Thr Lys Arg Ser Pro	
	35 40 45	
	Leu Thr Arg Ala His Leu Thr Glu Val Glu Ser Arg Leu Glu Arg Leu	
	50 55 60	
45	Glu Gln Leu Phe Leu Leu Ile Phe Pro Arg Glu Asp Leu Asp Met Ile	
	65 70 75 80	
	Leu Lys Met Asp Ser Leu Gln Asp Ile Lys Ala Leu Leu Thr Gly Leu	
	85 90 95	
50	Phe Val Gln Asp Asn Val Asn Lys Asp Ala Val Thr Asp Arg Leu Ala	
	100 105 110	
	Ser Val Glu Thr Asp Met Pro Leu Thr Leu Arg Gln His Arg Ile Ser	
	115 120 125	

Ala Thr Ser Ser Ser Glu Glu Ser Ser Asn Lys Gly Gln Arg Gln Leu  
 130 135 140 145

5 Thr Val Ser  
 145

<210> 43

<211> 606

10 <212> DNA

<213> Artificial Sequence

<220>

<221> misc\_feature

15 <223> Novel Sequence

<400> 43

20	atgaaagcgt taacggccag gcaacaagag gtgtttgate tcatccgtga tcacatcagc	60
	cagacaggta tgccgcgcac gcgtycgaa atcgccgcgc gtttgggtt ccgttccccca	120
	aacgcggctg aagaacatct gaaggcgctg gcacgcaaag gcgttattga aattgttccc	180
	ggcgcatac gcgggatctg tctgttgag gaagaggaag aagggttgcg gctggtaggt	240
25	cgttggctg ccggtaacc acttctggcg caacagcata ttgaaggtca ttatcaggtc	300
	gatcttccttattcaagcc gaatgtctat ttccctgtgc gggtcagcgg gatgtcgatg	360
	aaagatatacg gatatatggaa tggtacttg ctggcagtgc ataaaaactca ggatgtacgt	420
30	aacggtcagg tctgtgtcgc acgttattgtat gacgaaggta ccgttaacgcg cctgaaaaaa	480
	caggccaata aagtcgaact gttgccagaa aatagcgcagt ttaaaaccaat tgcgttagat	540
	cttcgtcagc agagcttccat tattaaaggcttggccggatgg gggtttatccg caacggcgcac	600
	tggctg	606

35 <210> 44.

<211> 202

<212> PRT

<213> Artificial Sequence

40 <220>

<221> misc feature

<223> Novel Sequence

<400> 44

45	Met Lys Ala Leu Thr Ala Arg Gln Gln Glu Val Phe Asp Leu Ile Arg	
	1 5 10 15	
	Asp His Ile Ser Gln Thr Gly Met Pro Pro Thr Arg Ala Glu Ile Ala	
	20 25 30	
50	Gln Arg Leu Gly Phe Arg Ser Pro Asn Ala Ala Glu Glu His Leu Lys	
	35 40 45	
	Ala Leu Ala Arg Lys Gly Val Ile Glu Ile Val Ser Gly Ala Ser Arg	
	50 55 60	

55

Gly Ile Arg Leu Leu Gln Glu Glu Glu Gly Leu Pro Leu Val Gly  
 65 70 75 80  
 5 Arg Val Ala Ala Gly Glu Pro Leu Leu Ala Gln Gln His Ile Glu Gly  
 85 90 95  
 His Tyr Gln Val Asp Pro Ser Leu Phe Lys Pro Asn Ala Asp Phe Leu  
 100 105 110  
 10 Leu Arg Val Ser Gly Met Ser Met Lys Asp Ile Gly Ile Met Asp Gly  
 115 120 125  
 Asp Leu Leu Ala Val His Lys Thr Gln Asp Val Arg Asn Gly Gln Val  
 130 135 140  
 15 Val Val Ala Arg Ile Asp Asp Glu Val Thr Val Lys Arg Leu Lys Lys  
 145 150 155 160  
 Gln Gly Asn Lys Val Glu Leu Leu Pro Glu Asn Ser Glu Phe Lys Pro  
 165 170 175  
 20 Ile Val Val Asp Leu Arg Gln Gln Ser Phe Thr Ile Glu Gly Leu Ala  
 180 185 190  
 Val Gly Val Ile Arg Asn Gly Asp Trp Leu  
 195 200

<210> 45  
 <211> 271  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <221> misc\_feature  
 30 <223> Novel Sequence  
 <400> 45

atggcccta aaaaagaagcg taaagtgcgc ccccccacccg atgtcagccct gggggacgag 60  
 35 ctccacttag acggcgagga cgtggcgatg gcgcattgcgc acgcgcgtaga cgatttcgtat 120  
 ctggacatgt tgggggacgg ggatccccg gggccggat ttaccccca cgaactccccc 180  
 ccctacggcg ctctggatat ggccgacttc gagtttgagc agatgttac cgatgcctt 240  
 40 ggaattgacg agtacggtgtt ggaattcccg g 271

<210> 46  
 <211> 90  
 <212> PRT  
 <213> Artificial Sequence  
 <220>  
 <221> misc feature  
 <223> Novel Sequence  
 50 <400> 46

Met Gly Pro Lys Lys Lys Arg Lys Val Ala Pro Pro Thr Asp Val Ser  
 1 5 10 15  
 55 Leu Gly Asp Glu Leu His Leu Asp Gly Glu Asp Val Ala Met Ala His

20 25 30

Ala	Asp	Ala	Leu	Asp	Asp	Phe	Asp	Leu	Asp	Met	Leu	Gly	Asp	Gly	Asp	
35						40						45				
5	Ser	Pro	Gly	Pro	Gly	Phe	Thr	Pro	His	Asp	Ser	Ala	Pro	Tyr	Gly	Ala
	50					55						60				
10	Leu	Asp	Met	Ala	Asp	Phe	Glu	Phe	Glu	Gln	Met	Phe	Thr	Asp	Ala	Leu
	65					70					75					80
	Gly	Ile	Asp	Glu	Tyr	Gly	Gly	Glu	Phe	Pro						
						85					90					

&lt;210&gt; 47

&lt;211&gt; 19

&lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;221&gt; misc feature

20 &lt;223&gt; Novel Sequence

&lt;400&gt; 47

ggagtlactgt cctccgagc 19

&lt;210&gt; 48

&lt;211&gt; 666

25 &lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;221&gt; misc feature

30 &lt;223&gt; Novel Sequence

&lt;400&gt; 48

ggatccccag cttggAACttc gagacggttat cagcaacaac acagtcatat ccatttcCAA 60

35 ttAGCTCTAC cacAGTGTGt gaaccATgt atccAGCACC acCTGtaACC aaaacaATT 120

tagaaGTAcTT tcacTTTGt aactGAGCTG tcattttatAT tGAATTTCa aaaATTCtta 180

ctttttttt ggatGGACGC aaAGAAGTT aataATCATA ttACATGGCA ttACCCACCAT 240

40 atACATATCC atATACATAT CCATATCTAA tCTTACACTCG ACTGCTGTAT ATAAAACCCAG 300

.tgGTTATATG tacAGTACTG CTGTATATAA aAccAGTGT tatATGtACa GtACGTCGAC 360

tgCTGTATAT AAAACCGATG GTTATATGTA cAGTACTGtCt GtATAaaaa ccAGTGGTTA 420

45 tATGtACAGT acGTCGAGGG AtGATAATGc gATTAGTTT ttAGCCTTAt ttCTGGGGTA 480

attaATCAGC GAAGCGATGA tTTTGTATCT AtTAACAGAT AtATAAAATGc AAAAACtGcA 540

taACCACTTT AACTAATACT ttCAACATT CGGTTGTA ttACTTTCTA ttCAAAATGTA 600

ataAAAGTAT caACAAAAAA ttGTTAATAT ACCTCTATAC ttAAcGtCA aggAGAAAAAA 660

50 actATA 666

&lt;210&gt; 49

&lt;211&gt; 1542

55 &lt;212&gt; DNA

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;221&gt; misc feature

&lt;223&gt; Novel Sequence

&lt;400&gt; 49

5	ctggacctga aacacgaagt ggcttaccga ggggtgtcc caggccaggt gaaggccaa	60
	ccgggggtcc aacaacggca ggtcaacggc cacgtgaggg actggatggc aggcggcgct	120
	ggtccaatt cgccgttcc gggagcggtg gtcacaaaaa acgcttaacaa tgggttttcg	180
10	tcggccactct ctcggaaat ctacggccc tacagtccaa atggaaaat aggccgttag	240
	gaactgtcgcc caggttcaag tataaatggg tgcatgtacag atggcgaggc acggacgtcg	300
	aagaaggggc ctggcccccc tcagcaagag gaactgtgtc tggtatgcgg ggacagagcc	360
15	tccggatacc actacaatgc gtcacgtgt gaagggtgt aagggttctt cagacggagt	420
	gttacaaaaa atgcgggttta tatttgtaaa ttccgtcacg cttcgcaaat ggacatgtac	480
	atgcgacgg aatgccagga gtgcgcctg aagaagtct tagctgttag catgaggct	540
20	gagtgcgtatc taccggagac tcagtgcgc atgaagcggaa aagagaagaa agcacaagaa	600
	gagaaggaca aactgcgtt cagcacgacg acgggtggacg accacatgcc gccttattatg	660
	cagtgtgaac ctccacccccc tgaagcgcga aggatttcacg aagtggtccc aaggtttctc	720
25	tccgacaagc tgttggagac aaacccggcgg aaaaacatcc cccaggatgc agccaaaccag	780
	cagtcccta tcggccaggtt catctgttac caggacgggtt acggacggcc ttctgtatgg	840
	gatttgaaga ggatttacgcgac gacgtggcag caagcggacg atgaaaacgcg agagtctgc	900
30	actcccttcc gccagatcac agagatgact atccctacgg tccaaacttat cgtggagttc	960
	gcgaagggtat tgccagggtt cgccaaagatc tcgcggctgtt atcaaattac gtgtttaag	1020
	gtctgtctaa gtggatataat gatgtcccgat gtcgcgcac gatacgatgc ggccctcagac	1080
35	agtgttctgtt tcggcaacaa ccaacgcgtt actcgcgcaca actaccggcaaa ggctggatgt	1140
	gcctacgtca tcggaggatct actgcacttc tgccgggtca tgtaactatggatggatgg	1200
	aacatccatt acgcgcgtct cacggctgtc gtcatttttt ctgaccggcc aggggttggag	1260
40	cagccgcacaa tgggtggaa aatccagcggt tactacactga atacgtcccg catctatata	1320
	ctgaaccacgc tgagcgggtc ggccgttcc tccgtcatat acggcaagat cctctcaatc	1380
	ctctctgtggc tacgcacgtt cggcatgcggaa aactccaaaca tggatcttc cctcaagctc	1440
45	aagaacagaa agtgcggcc ttctctcgag gagatctggg atgtggcgaa catgtcgac	1500
	acccaacccgc cgcctatccctt cgagttcccccc acgaatctctt ag	1542

&lt;210&gt; 50

&lt;211&gt; 513

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;221&gt; misc feature

&lt;223&gt; Novel Sequence

&lt;400&gt; 50

55

Leu Asp Leu Lys His Glu Val Ala Tyr Arg Gly Val Leu Pro Gly Gln  
 1 5 10 15

5 Val Lys Ala Glu Pro Gly Val His Asn Gly Gln Val Asn Gly His Val  
 20 25 30

Arg Asp Trp Met Ala Gly Gly Ala Gly Ala Asn Ser Pro Ser Pro Gly  
 35 40 45

10 Ala Val Ala Gln Pro Gln Pro Asn Asn Gly Tyr Ser Ser Pro Leu Ser  
 50 55 60

Ser Gly Ser Tyr Gly Pro Tyr Ser Pro Asn Gly Lys Ile Gly Arg Glu  
 65 70 75 80

15 Glu Leu Ser Pro Ala Ser Ser Ile Asn Gly Cys Ser Thr Asp Gly Glu  
 85 90 95

Ala Arg Arg Gln Lys Lys Gly Pro Ala Pro Arg Gln Gln Glu Leu  
 100 105 110

20 Cys Leu Val Cys Gly Asp Arg Ala Ser Gly Tyr His Tyr Asn Ala Leu  
 115 120

Thr Cys Glu Gly Cys Lys Gly Phe Phe Arg Arg Ser Val Thr Lys Asn  
 130 135 140

25 Ala Val Tyr Ile Cys Lys Phe Gly His Ala Cys Glu Met Asp Met Tyr  
 145 150 155 160

Met Arg Arg Lys Cys Gln Glu Cys Arg Leu Lys Lys Cys Leu Ala Val  
 165 170 175

30 Gly Met Arg Pro Glu Cys Val Val Pro Glu Thr Gin Cys Ala Met Lys  
 180 185 190

Arg Lys Glu Lys Lys Ala Gln Lys Glu Lys Asp Lys Leu Pro Val Ser  
 195 200 205

35 Thr Thr Val Asp Asp His Met Pro Pro Ile Met Gln Cys Glu Pro  
 210 215 220

Pro Pro Pro Glu Ala Ala Arg Ile His Glu Val Val Pro Arg Phe Leu  
 225 230 235 240

Ser Asp Lys Leu Leu Glu Thr Asn Arg Gln Lys Asn Ile Pro Gln Leu  
 245 250 255

40 Thr Ala Asn Gln Gln Phe Leu Ile Ala Arg Leu Ile Trp Tyr Gln Asp  
 260 265 270

Gly Tyr Glu Gln Pro Ser Asp Glu Asp Leu Lys Arg Ile Thr Gln Thr  
 275 280 285

45 Trp Gln Gln Ala Asp Asp Glu Asn Glu Glu Ser Asp Thr Pro Phe Arg

	290	295	300	
	Gln Ile Thr Glu Met Thr	Ile Leu Thr Val	Gln Leu Ile Val Glu Phe	
5	305	310	315	320
	Ala Lys Gly Leu Pro Gly Phe Ala Lys	Ile Ser Gln Pro Asp	Gln Ile	
	325	330	335	
	Thr Leu Leu Lys Ala Cys Ser Ser Glu Val Met Met Leu Arg Val Ala			
10	340	345	350	
	Arg Arg Tyr Asp Ala Ala Ser Asp Ser Val Leu Phe Ala Asn Asn Gln			
	355	360	365	
	Ala Tyr Thr Arg Asp Asn Tyr Arg Lys Ala Gly Met Ala Tyr Val Ile			
15	370	375	380	
	Glu Asp Leu Leu His Phe Cys Arg Cys Met Tyr Ser Met Ala Leu Asp			
	385	390	395	400
	Asn Ile His Tyr Ala Leu Leu Thr Ala Val Val Ile Phe Ser Asp Arg			
20	405	410	415	
	Pro Gly Leu Glu Gln Pro Gln Leu Val Glu Glu Ile Gln Arg Tyr Tyr			
	420	425	430	
	Leu Asn Thr Leu Arg Ile Tyr Ile Leu Asn Gln Leu Ser Gly Ser Ala			
	435	440	445	
25	Arg Ser Ser Val Ile Tyr Gly Lys Ile Leu Ser Ile Leu Ser Glu Leu			
	450	455	460	
	Arg Thr Leu Gly Met Gln Asn Ser Asn Met Cys Ile Ser Leu Lys Leu			
	465	470	475	480
30	Lys Asn Arg Lys Leu Pro Pro Phe Leu Glu Glu Ile Trp Asp Val Ala			
	485	490	495	
	Asp Met Ser His Thr Gln Pro Pro Pro Ile Leu Glu Ser Pro Thr Asn			
	500	505	510	

## 35            Leu

	<210> 51	
	<211> 4375	
	<212> DNA	
40	<213> Artificial Sequence	
	<220>	
	<221> misc_feature	
	<223> Novel Sequence	
	<400> 51	
45	tgtatttg atggcgccg tcatgcacgg tggccatata tgccatccag tcgaataga	60
	aaaaaaaaaa aaaaaaaaaat atcagttgtt ttgtccctcg ctgcgtttcg agtgttatcg	120
50	aatattttaga cgtcataatt cacgactgtc tttaaattt atatacgat taacggggcc	180
	gtttgttggc cgtgcgttg cgtttagtgg agtgcaggga tagtggccg agtatgttag	240
	ttcgtgtca tgtcaagtgt ggcgaagaaa gacaagccga cgatgtcggt gacggcgctg	300

	atcaaactggg	cgcggccggc	gccggcaggc	ccggccagc	cgca	cgactc	gtcg	ccctcg	cg	360	
5	ccggcagcca	tgctgc	agca	gctccc	gac	cagt	caat	gc	atgc	420	
	actgtcgatt	gctc	gctcg	tatgc	agt	cgat	ttaa	ccacat	ccca	480	
	atgtcacctc	ctg	agatgaa	acc	ca	gac	acc	ccat	gtcg	540	
	acttcgccc	cta	aacttca	ga	act	acc	cct	atcg	atgg	600	
10	ctatgctta	ctat	gccc	cg	gg	ggact	atgg	gt	tg	660	
	ggatgcaagg	gtt	ttca	gg	ggac	cc	ggagg	acc	tg	720	
	gaggcgaact	gcat	ata	ca	agg	gaca	agg	aaec	gt	780	
	aagt	ttt	gg	tt	gg	ggat	gg	ccat	at	840	
	cgccgc	cg	gg	gg	gg	ggat	gg	ggat	tc	900	
15	gagccctaa	cg	gagat	gg	gt	ttt	gg	tc	tt	960	
	cgcg	gg	gg	gg	gg	at	gg	gg	cc	1020	
20	caaataggca	aca	agca	aa	at	gccc	gt	cc	ct	1080	
	ggc	ag	ctgg	at	gg	at	gg	cc	at	1140	
25	ctttcgcca	tc	gct	gg	tc	tat	ttgg	at	gg	1200	
	ggaa	ac	gc	cc	act	tt	gg	cc	at	1260	
	cgca	act	cg	c	gg	at	tt	cc	gt	1320	
30	agtc	taa	gaga	tc	at	gg	ttt	gg	cc	1380	
	gt	gt	gt	at	tt	at	gg	gg	at	1440	
	aaaaaaatgt	tct	tttgc	ct	gg	ac	act	tc	tttgc	1500	
35	cg	tttgc	gt	cttgc	gt	cc	tttgc	ct	atc	1560	
	gaac	ac	cct	c	tc	tc	gg	cc	at	1620	
	gc	cttgc	cc	tc	ca	at	gg	cc	tg	1680	
40	gc	cttgc	cc	tc	ca	at	gg	cc	at	1740	
	ggactt	act	tttgc	at	tttgc	at	tttgc	at	tttgc	1800	
	gtc	at	aa	tttata	aa	ca	tttata	ac	tttata	1860	
45	tg	tttgc	at	tttgc	at	tttgc	at	tttgc	at	1920	
	aag	tttgc	at	tttgc	at	tttgc	at	tttgc	at	1980	
	atta	ttt	tttata	tttata	tttata	tttata	tttata	tttata	tttata	2040	
50	tttt	tttgc	tttata	tttata	tttata	tttata	tttata	tttata	tttata	2100	
	tc	atc	aaat	tc	at	tttgc	at	tttgc	at	2160	
	tc	atc	aaat	tc	at	tttgc	at	tttgc	at	2220	
	gg	agg	caga	at	tttgc	ata	tttgc	at	tttgc	at	2280



ctatggttgg	gtatgaattc	tatgttggaa	gtgttcgcgc	ggttgttaaa	atgattata	4260
atgatgttca	ctaaatatct	gactaaatgt	aagttat	tttttgtata	gacatagctt	4320
5	taagatgaag	gtgattaaac	tttatcccta	tcacaataaa	aaaaaaaaaa	aaaaaa
						4375

<210> 52  
<211> 472  
10 <212> PRT  
<213> Artificial Sequence  
<220>  
<221> misc\_feature  
<223> Novel Sequence  
15 <400> 52

Met Ser Ser Val Ala Lys Lys Asp Lys Pro Thr Met Ser Val Thr Ala						
1	5	10	15			
Leu Ile Asn Trp Ala Arg Pro Ala Pro Pro Gly Pro Pro Gln Pro Gln						
20	20	25	30			
Ser Ala Ser Pro Ala Pro Ala Ala Met Leu Gln Gln Leu Pro Thr Gln						
35	35	40	45			
Ser Met Gln Ser Leu Asn His Ile Pro Thr Val Asp Cys Ser Leu Asp						
50	50	55	60			
Met Gln Trp Leu Asn Leu Glu Pro Gly Phe Met Ser Pro Met Ser Pro						
65	65	70	75	80		
Pro Glu Met Lys Pro Asp Thr Ala Met Leu Asp Gly Leu Arg Asp Asp						
80	85	90	95			
Ala Thr Ser Pro Pro Asn Phe Lys Asn Tyr Pro Pro Asn His Pro Leu						
100	100	105	110			
Ser Gly Ser Lys His Leu Cys Ser Ile Cys Gly Asp Arg Ala Ser Gly						
115	115	120	125			
Lys His Tyr Gly Val Tyr Ser Cys Glu Gly Cys Lys Gly Phe Phe Lys						
130	130	135	140			
Arg Thr Val Arg Lys Asp Leu Ser Tyr Ala Cys Arg Glu Glu Arg Asn						
145	145	150	155	160		
Cys Ile Ile Asp Lys Arg Gln Arg Asn Arg Cys Gln Tyr Cys Arg Tyr						
165	165	170	175			
Gln Lys Cys Leu Ala Cys Gly Met Lys Arg Glu Ala Val Gln Glu Glu						
180	180	185	190			
Arg Gln Arg Asn Ala Arg Gly Ala Glu Asp Ala His Pro Ser Ser Ser						
195	195	200	205			
Val Gln Val Ser Asp Glu Leu Ser Ile Glu Arg Leu Thr Glu Met Glu						
210	210	215	220			
Ser Leu Val Ala Asp Pro Ser Glu Glu Phe Gln Phe Leu Arg Val Gly						
225	225	230	235	240		
Pro Asp Ser Asn Val Pro Pro Arg Tyr Arg Ala Pro Val Ser Ser Leu						

	245	250	255
	Cys Gln Ile Gly Asn Lys Gln Ile Ala Ala Leu Val Val Trp Ala Arg		
5	260 265 270		
	Asp Ile Pro His Phe Gly Gln Leu Glu Leu Asp Asp Gln Val Val Leu		
	275 280 285		
	Ile Lys Ala Ser Trp Asn Glu Leu Leu Leu Phe Ala Ile Ala Trp Arg		
10	290 295 300		
	Ser Met Glu Tyr Leu Glu Asp Glu Arg Glu Asn Gly Asp Gly Thr Arg		
	305 310 315 320		
	Ser Thr Thr Gln Pro Gln Leu Met Cys Leu Met Pro Gly Met Thr Leu		
15	325 330 335		
	His Arg Asn Ser Ala Gln Gln Ala Gly Val Gly Ala Ile Phe Asp Arg		
	340 . 345 350		
	Val Leu Ser Glu Leu Ser Leu Lys Met Arg Thr Leu Arg Met Asp Gln		
	355 360 365		
20	Ala Glu Tyr Val Ala Leu Lys Ala Ile Val Leu Leu Asn Pro Asp Val		
	370 375 380		
	Lys Gly Leu Lys Asn Arg Gln Glu Val Asp Val Leu Arg Glu Lys Met		
	385 390 395 400		
25	Phe Ser Cys Leu Asp Asp Tyr Cys Arg Arg Ser Arg Ser Asn Glu Glu		
	405 410 415		
	Gly Arg Phe Ala Ser Leu Leu Leu Arg Leu Pro Ala Leu Arg Ser Ile		
	420 425 430		
30	Ser Leu Lys Ser Phe Glu His Leu Tyr Phe Phe His Leu Val Ala Glu		
	435 440 445		
	Gly Ser Ile Ser Gly Tyr Ile Arg Glu Ala Leu Arg Asn His Ala Pro		
	450 455 460		
35	Pro Ile Asp Val Asn Ala Met Met		
	465 470		

<210> 53  
 <211> 1404  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <221> misc feature  
 <223> Novel Sequence  
 <400> 53

	atggacacca aacatttcct gccgctcgac ttcttaccc aggtgaactc ttctgcctc	60
50	aactctcaa cgggtcgagg ctccatggct gtccccctgc tgccacccctc ctgggtccg	120
	ggaatcggt ctccactggg ctccgctggg cagctgcact ctccatcatcg cacccctgagc	180
	tccccatca atggcatggg tccggcccttc tctgtcatca gctcccccata gggcccgac	240
	tccatgtcggt taccacccac acccacattg ggcttcggga ctggtagccc ccagctaat	300

55

	tcacccatga accctgttag cagcaactgag gatatacaggc cgccactagg cctcaatggc	360
5	gttctcaagg ttccctgcca tccctcaggaa aatatggcct cttccaccaa gcacatctgt	420
	gctatctgtg gggaccgcctc ctcaggcaaa cactatgggg tatacagtt tgagggtgc	480
	aaggggctt tcaagaggac agtacgcaaa gacctgaccc acacctgccc agacaacaag	540
	gactgcctga tcgacacaagag acagcggAAC cggtgtcagt actgccccta ccagaagtg	600
10	ctggccatgg gcatagaagcg ggaagctgtg caggaggagc ggcagccccgg caaggaccgg	660
	aatgagaacg aggtggggatc caccacqagt gccaacaggAC acatgcgtg agagaagatt	720
	ctggaaAGCC agttgtgtg cggacccaaAG actgagacat acgtggggc aaacatgggg	780
15	ctgaacccca gtcacccaa tgaccctgtt accaacatct gtcacggc agacaaggcag	840
	ctcttcactc ttgtggagtG ggccaagagg atccccacact tttctgagct gcccctagac	900
	gaccaggctca tcctgtctacg ggcaggctgg aacgagctgc tgatgcctc ctctcccac	960
20	cgctccatag ctgtgaaaga tgggatctc ctggccacccg gcctgcacgt acaccggAAC	1020
	agcgctcaca gtgtctgggtt gggcgccate tttgacaggG tgcttaacaga gctgggtct	1080
	aagatgcgtg acatgcgat ggacaaggAC gagctgggtt gcctgcgAGC cattgtccctg	1140
25	ttcaaccccty actctaaaggG gctctcaaac cctgtgtgggg tggaggcgtt gaggggaaag	1200
	gtgtatgcgt cactagaAGC gtactgcaaa cacaagtacc ctgagcagcc gggcaggTTT	1260
	gccaagctgc tgctccgcct gctgtcactg cgttccatcg ggctcaatgC cttggaggcac	1320
30	ctgtttttct tcaagctcat cggggacacg cccatcgaca ctttcctcat ggagatgtg	1380
	gaggaccacatc acatcgaccatc cttag	1404

&lt;210&gt; 54

&lt;211&gt; 467

35 &lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

&lt;220&gt;

&lt;221&gt; misc feature

&lt;223&gt; Novel Sequence

40 &lt;400&gt; 54

Met	Asp	Thr	Lys	His	Phe	Leu	Pro	Leu	Asp	Phe	Ser	Thr	Gln	Val	Asn	
1					5				10				15			
45	Ser	Ser	Ser	Leu	Asn	Ser	Pro	Thr	Gly	Arg	Gly	Ser	Met	Ala	Val	Pro
							20		25				30			
	Ser	Leu	His	Pro	Ser	Leu	Gly	Pro	Gly	Ile	Gly	Ser	Pro	Leu	Gly	Ser
						35		40				45				
50	Pro	Gly	Gln	Leu	His	Ser	Pro	Ile	Ser	Thr	Leu	Ser	Ser	Pro	Ile	Asn
						50		55			60					
	Gly	Met	Gly	Pro	Pro	Phe	Ser	Val	Ile	Ser	Ser	Pro	Met	Gly	Pro	His
						65		70			75			80		

Ser Met Ser Val Pro Thr Thr Pro Thr Leu Gly Phe Gly Thr Gly Ser  
 85 90 95  
 5 Pro Gln Leu Asn Ser Pro Met Asn Pro Val Ser Ser Thr Glu Asp Ile  
 100 105 110  
 Lys Pro Pro Leu Gly Leu Asn Gly Val Leu Lys Val Pro Ala His Pro  
 115 120 125  
 10 Ser Gly Asn Met Ala Ser Phe Thr Lys His Ile Cys Ala Ile Cys Gly  
 130 135 140  
 Asp Arg Ser Ser Gly Lys His Tyr Gly Val Tyr Ser Cys Glu Gly Cys  
 145 150 155 160  
 15 Lys Gly Phe Phe Lys Arg Thr Val Arg Lys Asp Leu Thr Tyr Thr Cys  
 165 170 175  
 Arg Asp Asn Lys Asp Cys Leu Ile Asp Lys Arg Gln Arg Asn Arg Cys  
 180 185 190  
 20 Gln Tyr Cys Arg Tyr Gln Lys Cys Leu Ala Met Gly Met Lys Arg Glu  
 195 200 205  
 Ala Val Gln Glu Glu Arg Gln Arg Gly Lys Asp Arg Asn Glu Asn Glu  
 210 215 220  
 25 Val Glu Ser Thr Ser Ser Ala Asn Glu Asp Met Pro Val Glu Lys Ile  
 225 230 235 240  
 Leu Glu Ala Glu Leu Ala Val Glu Pro Lys Thr Glu Thr Tyr Val Glu  
 245 250 255  
 30 Ala Asn Met Gly Leu Asn Pro Ser Ser Pro Asn Asp Pro Val Thr Asn  
 260 265 270  
 Ile Cys Gln Ala Ala Asp Lys Gln Leu Phe Thr Leu Val Glu Trp Ala  
 275 280 285  
 35 Lys Arg Ile Pro His Phe Ser Glu Leu Pro Leu Asp Asp Gln Val Ile  
 290 295 300  
 Leu Leu Arg Ala Gly Trp Asn Glu Leu Leu Ile Ala Ser Phe Ser His  
 305 310 315 320  
 Arg Ser Ile Ala Val Lys Asp Gly Ile Leu Leu Ala Thr Gly Leu His  
 325 330 335  
 40 Val His Arg Asn Ser Ala His Ser Ala Gly Val Gly Ala Ile Phe Asp  
 340 345 350  
 Arg Val Leu Thr Glu Leu Val Ser Lys Met Arg Asp Met Gln Met Asp  
 355 360 365  
 45 Lys Thr Glu Leu Gly Cys Leu Arg Ala Ile Val Leu Phe Asn Pro Asp  
 370 375 380  
 Ser Lys Gly Leu Ser Asn Pro Ala Glu Val Glu Ala Leu Arg Glu Lys  
 385 390 395 400  
 50 Val Tyr Ala Ser Leu Glu Ala Tyr Cys Lys His Lys Tyr Pro Glu Gln  
 405 410 415  
 Pro Gly Arg Phe Ala Lys Leu Leu Leu Arg Leu Pro Ala Leu Arg Ser

	420	425	430
5	Ile Gly Leu Lys Cys' Leu Glu His Leu Phe Phe Phe Lys Leu Ile Gly 435 440 445		
	Asp Thr Pro Ile Asp Thr Phe Leu Met Glu Met Leu Glu Ala Pro His 450 455 460		
10	Gln Ala Thr 465		
	<210> 55		
	<211> 309		
	<212> DNA		
15	<213> Artificial Sequence		
	<220>		
	<221> misc_feature		
	<223> Novel Sequence		
	<400> 55		
20	ggtgtggaaa gtccccaggc tccccagcac gcagaagttat gcaagcatg catctcaatt agtccacaa cagggtgtgg aagtccccag gtcctccaggc aggcagaagt atgcaaaagca tgcatctcaa tttagtcagca accatagttcc cggccctaacc tccggccatc cggcccttaa ctccggcccg ttccggccat tctccggccc atggctgact aattttttttt atttatgcag aggccgaggc cgccctggcc tctgagctat tccagaagta gtgaggaggc tttttggag gccttaggct	60 120 180 240 300 309	
	<210> 56		
	<211> 24		
	<212> DNA		
35	<213> Artificial Sequence		
	<220>		
	<221> misc feature		
	<223> Novel-Sequence		
	<400> 56		
40	tataataatgg atccccgggt acccg 24		
	<210> 57		
	<211> 1653		
	<212> DNA		
45	<213> Artificial Sequence		
	<220>		
	<221> misc feature		
	<223> Novel Sequence		
	<400> 57		
50	atggaaagacg ccaaaaaacat aaagaaaaggc cccggcccat tctatctct agaggatgga accggctggag agcaactgca taaggtatg aagagatacg ccctggttcc tggaaacaatt gctttacag atgcacatata cgaggtgaac atcacgtacg cgaaatactt cgaaatgtcc	60 120 180	

<210> 58  
<211> 867  
<212> DNA  
<213> Artificial Sequence  
<220>  
<221> misc feature  
<223> Novel Sequence  
<400> 58

	ccgatgtact cggtagcaggta aagcgatgag ctgtcaatcg agcgcctaacc ggagatggag	120
5	tcttggtgg cagatcccag cgaggaggtc cagttctcc gcgtggggcc tgacagcaac	180
	gtgcctccac gtatccgcgc gccccctcc tccctctgcc aaataggcaaa caagcaata	240
	ggcggttgg tggatgggc ggcgcacate cctcatttcg ggcagctgg gctggacgat	300
10	caagtggta tcatacaaggc ctcctggaaat gagctgtac tcttcggcat cgcctggcgc	360
	tctatggagt atttggaaaga tgagagggag aacggggacg gaacgcggag caccactcg	420
	ccacaactga tggtgtctcat gcctggcatg acggtgcacc gcaactggcgc gcagcaggcg	480
15	ggcggtggcg ccatcttcga cgcgtgtcgtc tccgagactca gtctgaagat ggcacccgt	540
	cgcgtggacc agggcggatc cgtcgccgtc aaagecatcg tgcgtgtcaa ccctgtatgt	600
	aaaggactga agaatcgca agaagggttgc gttttgcggaaaaaaatgtt ctcttgctcg	660
	gacgactact gccggcggtc gcaagcaac gaggaaaggc ggtttgcgtc cttgtgtcg	720
20	cggtgtcccg ctctccgtc catctcggtc aagagtttcg aacacctata cttcttccac	780
	ctcggtggccg aagggtccat cagcggatac atacgagagg cgctccgaaa ccacgcgcct	840
	ccgatcgacg tcaatggcat gatgtaa	867

25  
<210> 59  
<211> 225  
<212> DNA  
<213> Artificial Sequence  
30 <220>  
<221> misc\_feature  
<223> Novel Sequence  
<400> 59

35	tcgacattgg acaagtgcata tgaacccttg tctctcgaga gacaaggggg ttcaatgcac	60
	ttgtccaatg tcgagagaca aggggggtca atgcacttgtt ccaatgtcgagacaagg	120
	gggtcaatgc actttgtccaa tgtcgagaga caagggggtt caatgcactt gtccaatgtc	180
40	gagagacaag ggggttcaat gcactttgtcc aatgtcgact ctaga	225

55 <210> 60  
<211> 619  
45 <212> DNA  
<213> Artificial Sequence  
<220>  
<221> misc feature  
<223> Novel Sequence  
50 <400> 60

	cgttacataa cttacggtaa atggccgcgc tggctgaccg cccaaacgacc cccgcgcatt	60
	gacgtcaata atgacgtatg ttccatagt aacgccaata gggactttcc attgacgtca	120
55	atgggtggag tatttacggtaaactggcca cttggcagta catcaagtgt atcatatgcc	180

	aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt atgcccagta	240
5	catgacccta tgggacttgc ctacttggca gtacatctac gttagtgatc tcgcattac	300
	catggtgatc cggttttggc agtacatcaa tggcggttgc tagcggtttg actcacgggg	360
	atttccaagt ctccacccca ttgacgtcaa tgggagtttgc tttggcacc aaaaatcaacg	420
	ggacttccaa aatgtcgta acaactccgc cccattgacg caaatggcg gtaggcgctgt	480.
10	acggggag gtctataataa gcagagctcg tttagtgaac cgtcagatcg cctggagacg	540
	ccatccacgc tgtttgcacc tccatagaag acacccggac cgatccagcc tccgcggccg	600
	ggaacgggtgc atttggaaacg	619
15	<210> 61	
	<211> 262	
	<212> DNA	
	<213> Artificial Sequence	
20	<220>	
	<221> misc_feature	
	<223> Novel Sequence	
	<400> 61	
25	atgtatgtttt atgcaataact ctgttagtct tgcaacatgg taacgatgg ttagcacat	60
	gccttacaag gagagaaaaa gcaccgtgca tgccgatagg tggaaatgg gtggatcgat	120
	cgtgccttat taggaaggca acagacgggt ctgacatggaa ttggacaaac cactgaattc	180
30	cgcatttcgac agatattgtt tttaaatgtcc tagtcgtata caataaacgc catttgcacca	240
	ttcacccat tggagtgacat	262
35	<210> 62	
	<211> 1247	
	<212> DNA	
	<213> Artificial Sequence	
40	<220>	
	<221> misc feature	
	<223> Novel Sequence	
	<400> 62	
45	tctattttctt caggccgtga ggaactgtcg ccagcttcaa gtataaatgg gtgcagttaca	60
	gatggcgagg caacgacgtca gaagaaggc cctgcgcggc gtcagacaaga ggaactgtgt	120
	ctggatgtcg gggacacagac ctcggatatacactacaatgt cgctcacgtg tgaagggtgt	180
	aaagggttctt tcagacggggatgttacaaa aatgcgggtt atatgttataa attcgggtcac	240
50	gcttgcggaaa tggacatgtt catgcgacgg aaatgccagg agtgcggcctt gaagaagtgc	300
	tttagctgtatgcatgaggcc tgagtgcgtatgcaccgaga ctcagtgccgc catgaacgg	360
	aaagagaaga aagcacacaa ggagaaggac aaactgcgttgcacgc gacgggtggac	420

	gaccacatgc cgccattat gcagtgtcaa cctccaccc tcgaagcagc aaggattcac	480
5	gaagtggtcc caaggtttct ctccgacaag ctgttgaga caaaccggca gaaaacatc	540
	ccccagttcaga cagccaacca gcagttccct atgcgcaggc tcatacgatca ccaggacggg	600
	tacgagcgc cttctgtatga agattgtaa aggattacgc agacgtggca gcaagcggac	660
10	gatgaaaacaa agagactgaa cactcccttc cgccagatca cagagatgc tatcctcagc	720
	gtccaaactta tcgttgagtt cgcgaaaggta ttgccagggt tcgccaagat ctcgcagcct	780
	gatcaaattt cgctgtttaa ggcttgctca agtggatgtatgcgtcc agtgcgcgca	840
15	cgatcgtatgc cggccctcaga cagtgatctgt ttcgcgaaca accaagcata cactcgcgcac	900
	eactaccgcg aggctggcat ggcctacgtc atcgaggatc tactgcattt ctgcgggtgc	960
	atgtactcta tggcggttgc caacatccat tacgcgcgtc tcacggctgt cgtcatcttt	1020
	tctgaccggc cagggttggc gcagccgaa ctgggtggaa aatccagcg gtactacctg	1080
20	aatacgtcc gcatatatat cctgaaccag ctgagccgggt cggcggttc gtccgtcata	1140
	tacggcaaga tcctctcaat cctctcttagt ctacgcacgc tcggcatgc aaactcoaac	1200
	atgtgcatactt ccctcaagct caagaacaga aagctggcgc ctttctt	1247

25

&lt;210&gt; 63

&lt;211&gt; 440

&lt;212&gt; PRT

&lt;213&gt; Artificial Sequence

30

&lt;220&gt;

&lt;221&gt; misc feature

&lt;223&gt; Novel Sequence

&lt;400&gt; 63

35

Ser	Ile	Ser	Ser	Gly	Arg	Glu	Glu	Leu	Ser	Pro	Ala	Ser	Ser	Ile	Asn
1				5				10						15	
Gly	Cys	Ser	Thr	Asp	Gly	Glu	Ala	Arg	Arg	Gln	Lys	Lys	Gly	Pro	Ala
				20				25						30	
Pro	Arg	Gln	Gln	Glu	Glu	Leu	Cys	Leu	Val	Cys	Gly	Asp	Arg	Ala	Ser
				35				40						45	
Gly	Tyr	His	Tyr	Asn	Ala	Leu	Thr	Cys	Glu	Gly	Cys	Lys	Gly	Phe	Phe
				50				55						60	
Arg	Arg	Ser	Val	Thr	Lys	Asn	Ala	Val	Tyr	Ile	Cys	Lys	Phe	Gly	His
				65				70						75	80
Ala	Cys	Glu	Met	Asp	Met	Tyr	Met	Arg	Arg	Lys	Cys	Gln	Cys	Arg	
				85				90						95	
Leu	Lys	Lys	Cys	Leu	Ala	Val	Gly	Met	Arg	Pro	Glu	Cys	Val	Val	Pro
				100										105	110
Glu	Thr	Gln	Cys	Ala	Met	Lys	Arg	Lys	Glu	Lys	Lys	Ala	Gln	Lys	Glu
				115				120						125	

55

Lys Asp Lys Leu Pro Val Ser Thr Thr Thr Val Asp Asp His Met Pro  
 130 135 140  
 Pro Ile Met Gln Cys Glu Pro Pro Pro Glu Ala Ala Arg Ile His  
 145 150 155 160  
 Glu Val Val Pro Arg Phe Leu Ser Asp Lys Leu Leu Glu Thr Asn Arg  
 165 170 175  
 Gln Lys Asn Ile Pro Gln Leu Thr Ala Asn Gln Gln Phe Leu Ile Ala  
 180 185 190  
 Arg Leu Ile Trp Tyr Gln Asp Gly Tyr Glu Gln Pro Ser Asp Glu Asp  
 195 200 205  
 Leu Lys Arg Ile Thr Gln Thr Trp Gln Gln Ala Asp Asp Glu Asn Gln  
 210 215 220  
 Glu Ser Asp Thr Pro Phe Arg Gln Ile Thr Glu Met Thr Ile Leu Thr  
 225 230 235 240  
 Val Gln Leu Ile Val Glu Phe Ala Lys Gly Leu Pro Gly Phe Ala Lys  
 245 250 255  
 Ile Ser Gln Pro Asp Gln Ile Thr Leu Leu Lys Ala Cys Ser Ser Glu  
 260 265 270  
 Val Met Met Leu Arg Val Ala Arg Arg Tyr Asp Ala Ala Ser Asp Ser  
 275 280 285  
 Val Leu Phe Ala Asn Asn Gln Ala Tyr Thr Arg Asp Asn Tyr Arg Lys  
 290 295 300  
 Ala Gly Met Ala Tyr Val Ile Glu Asp Leu Leu His Phe Cys Arg Cys  
 305 310 315 320  
 Met Tyr Ser Met Ala Leu Asp Asn Ile His Tyr Ala Leu Leu Thr Ala  
 325 330 335  
 Val Val Ile Phe Ser Asp Arg Pro Gly Leu Glu Gln Pro Gln Leu Val  
 340 345 350  
 Glu Glu Ile Gln Arg Tyr Tyr Leu Asn Thr Leu Arg Ile Tyr Ile Leu  
 355 360 365  
 Asn Gln Leu Ser Gly Ser Ala Arg Ser Ser Val Ile Tyr Gly Lys Ile  
 370 375 380  
 Leu Ser Ile Leu Ser Glu Leu Arg Thr Leu Gly Met Gln Asn Ser Asn  
 385 390 395 400  
 Met Cys Ile Ser Leu Lys Leu Lys Asn Arg Lys Leu Pro Pro Phe Leu  
 405 410 415  
 Glu Glu Ile Trp Asp Val Ala Asp Met Ser His Thr Gln Pro Pro Pro  
 420 425 430  
 Ile Leu Glu Ser Pro Thr Asn Leu  
 435 440

<210> 64  
 <211> 943  
 <212> DNA  
 <213> Artificial Sequence  
 <220>  
 <221> misc feature  
 <223> Novel Sequence  
 <400> 64

	atgacttcga aagtttatga tccagaacaa aggaaacggg tgataactgg tccgcagtgg	60
5	tggggcagat gtaacaatat gaatgttctt gattcatat ttaattatta tgattcagaa	120
	aaacatgcgg aaaatgtgt tatTTTTTAA catggtaacg cggccttc ttatttatgg	180
	cgacatgttg tgccacatat tgagccagta gcgcgggtta ttataccaga ctttttttgtt	240
	atggcataat caggcaaatc tggtatggt tcTTTATGGT tacttgatca ttacaatat	300
10	cTTTACTGcat ggTTTGAact tcTTTAAATTA ccaaAGAAGA tcTTTTGTG cggccatgt	360
	tgggggtctt gtttgcatt tcTTTATGGT tacttgatca aagataagat caaAGCAATA	420
	gttcaCGCTG aaAGTGTAGT agatgttgat gaatcatggg atgaatGCC tgatATTGAA	480
15	gaagatATTG cgttgcataa atctgaagaa ggagaaaaaa tggttttggaa gaataacttc	540
	ttcgtggaaa ccatttttgcC atcaaaaaatc atgaaaaAGT tagaaccaga agaatttgcA	600
	gcataCTTG aaccattCAA agagaaAGGT gaagttcgtc gtccaaCATT atcatggcct	660
20	cgtgaaATCC cgttagtaaa aggtggtaaa CCTGACGTG tacaatTTGT taggaattt	720
	aatgttttac tacgtgcAAG tgatgatTTA ccaaaATGT ttattgtatcc ggacCCAGGA	780
	ttcTTTTCCA atgctattgt tgaaaggTGC aagaaggTTc ctaataCTGA atttGTcaAA	840
25	gtaaaaAGGTc ttCATTTTC GCAAGAAGAT GCACTGTG AAATGGAAA ATATATCAA	900
	tcgTTTGTG agcgagttct caaaaaaaaaAGAAGATTTCT AGA	943

## 30 Claims

## 1. A gene expression modulation system comprising:

35 a) a first gene expression cassette that is capable of being expressed in a host cell comprising a polynucleotide sequence that encodes a first polypeptide comprising:

- i) a DNA-binding domain that recognizes a response element associated with a gene whose expression is to be modulated;
- ii) a ligand binding domain comprising a ligand binding domain from an ecdysone receptor polypeptide;

40 b) a second gene expression cassette that is capable of being expressed in the host cell comprising a polynucleotide sequence that encodes a second polypeptide comprising:

- i) a transactivation domain; and
- ii) a ligand binding domain comprising a ligand binding domain from a truncated retinoid X receptor comprising a deletion of at least the A and B domains;

50 wherein the transactivation domain is from a nuclear receptor other than an ecdysone receptor, a retinoid X receptor, or an ultraspirelre receptor, and wherein the ligand binding domains from the first polypeptide and the second polypeptide are different and dimerize.

## 2. The gene expression modulation system according to claim 1, further comprising a third gene expression cassette comprising:

- i) a response element to which the DNA-binding domain of the polypeptide binds;
- ii) a promoter that is activated by the transactivation domain of the second polypeptide; and
- iii) the gene whose expression is to be modulated.

3. The gene expression modulation system according to claim 1, wherein the ligand binding domain of the first polypeptide is encoded by a polynucleotide comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SED ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, and SEQ ID NO: 10.

5. The gene expression modulation system according to claim 1, wherein the ligand binding domain of the first polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, and SEQ ID NO: 20.

10. The gene expression modulation system according to claim 1, wherein the ligand binding domain of the second polypeptide is encoded by a polynucleotide comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, and SEQ ID NO: 30.

15. The gene expression modulation system according to claim 1, wherein the ligand binding domain of the second polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO 37, SEQ ID NO 38, SEQ ID NO: 39, and SEQ ID NO: 40.

20. A gene expression modulation system comprising:

a) a first gene expression cassette that is capable of being expressed in a host cell comprising a polynucleotide sequence that encodes a first polypeptide comprising:

25 i) a DNA-binding domain that recognizes a response element associated with a gene whose expression is to be modulated; and  
ii) a ligand binding domain comprising a ligand binding domain from a truncated retinoid X receptor comprising a deletion of at least the A and B domains; and

30 b) a second gene expression cassette that is capable of being expressed in a host cell comprising a polynucleotide sequence that encodes a second polypeptide comprising:  
i) a transactivation domain; and  
ii) a ligand binding domain comprising a ligand binding domain from an ecdysone receptor;

35 wherein the ligand binding domains from the first polypeptide and the second polypeptide are different and dimerize.

40 8. The gene expression modulation system according to claim 7, further comprising a third gene expression cassette comprising:

i) a response element to which the DNA-binding domain of the first polypeptide binds;  
ii) a promoter that is activated by the transactivation domain of the second polypeptide; and  
iii) the gene whose expression is being modulated.

45 9. The gene expression modulation system according to claim 7, wherein the ligand binding domain of the first polypeptide is encoded by a polynucleotide comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, and SEQ ID NO: 30.

50 10. The gene expression modulation system according to claim 7, wherein the ligand binding domain of the first polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO 37, SEQ ID NO 38, SEQ ID NO: 39, and SEQ ID NO: 40.

55 11. The gene expression modulation system according to claim 7, wherein the ligand binding domain of the second polypeptide is encoded by a polynucleotide comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SED ID NO: 7, SEQ

ID NO: 8, SEQ ID NO: 9, and SEQ ID NO: 10.

12. The gene expression modulation system according to claim 7, wherein the ligand binding domain of the second polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NO 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, and SEQ ID NO: 20.

13. A gene expression cassette comprising a polynucleotide encoding a hybrid polypeptide selected from the group consisting of:

- 10 a) a DNA-binding domain and an ecdysone receptor ligand binding domain, wherein the DNA binding domain is from a nuclear receptor other than an ecdysone receptor;
- b) a DNA-binding domain and a ligand binding domain from a truncated retinoid X receptor comprising a deletion of at least the A and B domains, wherein the DNA-binding domain is from a nuclear receptor other than a retinoid X receptor;
- c) a transactivation domain and an ecdysone receptor ligand binding domain, wherein the transactivation domain is from a nuclear receptor other than an ecdysone receptor, and
- d) a transactivation domain and a ligand binding domain from a truncated retinoid X receptor comprising a deletion of at least the A and B domains, wherein the transactivation domain is from a nuclear receptor other than a retinoid X receptor.

14. The gene expression cassette according to claim 13, wherein said expression cassette comprises a polynucleotide encoding a hybrid polypeptide comprising:

- 25 a DNA-binding domain and an ecdysone receptor ligand binding domain wherein the DNA binding domain is from a nuclear receptor other than an ecdysone receptor; or
- a DNA-binding domain and a retinoid X receptor ligand binding domain, wherein the DNA-binding domain is from a nuclear receptor other than a retinoid X receptor; and

30 wherein the DNA-binding domain is GAL4 DNA-binding domain or a LexA DNA-binding domain.

15. The gene expression cassette according to claim 13, wherein said expression cassette comprises a polynucleotide encoding a hybrid polypeptide comprising:

- 35 a transactivation domain and an ecdysone receptor ligand binding domain, wherein the transactivation domain is from a nuclear receptor other than an ecdysone receptor;
- or a transactivation domain and a retinoid X receptor ligand binding domain, wherein the transactivation domain is from a nuclear receptor other than a retinoid X receptor;

40 and wherein the transactivation domain is a VP16 transactivation domain.

16. A gene expression cassette comprising a polynucleotide encoding a hybrid polypeptide comprising a DNA-binding domain encoded by a polynucleotide comprising a nucleic acid sequence selected from the group consisting of GAL4 DBD (SEQ ID NO: 41) or a LexA DBD (SEQ ID NO: 43) and an ecdysone receptor ligand binding domain encoded by a polynucleotide comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SED ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, and SEQ ID NO: 10.

50 17. A gene expression cassette comprising a polynucleotide encoding a hybrid polypeptide comprising a DNA-binding domain comprising an amino acid sequence selected from the group consisting of a GAL4 DBD (SEQ ID NO: 42) or a LexA DBD (SEQ ID NO: 44) and an ecdysone receptor ligand binding domain comprising an amino acid sequence selected from the group consisting of SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, and SEQ ID NO: 20.

55 18. A gene expression cassette comprising a polynucleotide encoding a hybrid polypeptide comprising a DNA-binding domain encoded by a polynucleotide comprising a nucleic acid sequence selected from the group consisting of a GAL4 DBD (SEQ ID NO: 41) or a LexA DBD (SEQ ID NO: 43) and a retinoid X receptor ligand binding domain encoded by a polynucleotide comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO:

21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, and SEQ ID NO: 30.

5 19. A gene expression cassette comprising a polynucleotide encoding a hybrid polypeptide comprising a DNA-binding domain comprising an amino acid sequence selected from the group consisting of a GAL4 DBD (SEQ ID NO: 42) or a LexA DBD (SEQ ID NO: 44) and a retinoid X receptor ligand binding domain comprising an amino acid sequence selected from the group consisting of SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO 37, SEQ ID NO 38, SEQ ID NO: 39, and SEQ ID NO: 40.

10 20. A gene expression cassette comprising a polynucleotide encoding a hybrid polypeptide comprising a transactivation domain encoded by a polynucleotide comprising a nucleic acid sequence of SEQ ID NO: 45 and an ecdysone receptor ligand binding domain encoded by a polynucleotide comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, and SEQ ID NO: 10.

15 21. A gene expression cassette comprising a polynucleotide encoding a hybrid polypeptide comprising a transactivation domain comprising an amino acid sequence of SEQ ID NO: 46 and an ecdysone receptor ligand binding domain comprising an amino acid sequence selected from the group consisting of SEQ ID NO: 11, SEQ ID NO: 12, SEQ ID NO: 13, SEQ ID NO: 14, SEQ ID NO: 15, SEQ ID NO: 16, SEQ ID NO: 17, SEQ ID NO: 18, SEQ ID NO: 19, and SEQ ID NO: 20.

20 22. A gene expression cassette comprising a polynucleotide encoding a hybrid polypeptide comprising a transactivation domain encoded by a polynucleotide comprising a nucleic acid sequence of SEQ ID NO: 45 and a retinoid X receptor ligand binding domain encoded by a polynucleotide comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, and SEQ ID NO: 30.

25 23. A gene expression cassette comprising a polynucleotide encoding a hybrid polypeptide comprising a transactivation domain comprising an amino acid sequence of SEQ ID NO: 46 and a retinoid X receptor ligand binding domain comprising an amino acid sequence selected from the group consisting of SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO 37, SEQ ID NO 38, SEQ ID NO: 39, and SEQ ID NO: 40.

30 24. An isolated polynucleotide encoding a truncated retinoid X receptor polypeptide wherein the polynucleotide comprises a nucleic acid sequence selected from the group consisting of SEQ ID NO: 21, SEQ ID NO: 22, SEQ ID NO: 23, SEQ ID NO: 24, SEQ ID NO: 25, SEQ ID NO: 26, SEQ ID NO: 27, SEQ ID NO: 28, SEQ ID NO: 29, and SEQ ID NO: 30.

35 25. An isolated truncated retinoid X receptor polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO: 31, SEQ ID NO: 32, SEQ ID NO: 33, SEQ ID NO: 34, SEQ ID NO: 35, SEQ ID NO: 36, SEQ ID NO 37, SEQ ID NO 38, SEQ ID NO: 39, and SEQ ID NO: 40.

40 26. An isolated polynucleotide comprising a polynucleotide sequence which encodes the isolated polypeptide according to claim 25.

45 27. A method of modulating the expression of a gene in a host cell comprising the gene to be modulated comprising the steps of:

- a) introducing into the host cell the gene expression modulation system according to claim 1 or 2; and
- b) introducing into the host cell a ligand that independently combines with the ligand binding domains of the first polypeptide and the second polypeptide;

wherein the gene to be expressed is a component of a chimeric gene comprising:

- i) a response element to which the DNA binding domain from the first polypeptide binds;
- ii) a promoter that is activated by the transactivation domain of the second polypeptide; and
- iii) a gene whose expression is to be modulated.

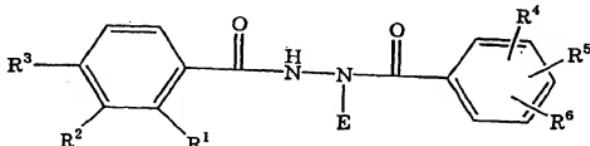
whereby a complex is formed comprising the ligand, the first polypeptide, and the second polypeptide, and whereby the complex modulates expression of the gene in the host cell.

28. The method according to claim 27, wherein the ligand is a compound of the formula:

5

10

15



wherein:

E is a ( $\text{C}_4\text{-C}_6$ )alkyl containing a tertiary carbon or a cyano( $\text{C}_3\text{-C}_5$ )alkyl containing a tertiary carbon:

20

$\text{R}^1$  is H, Me, Et, i-Pr, F, formyl,  $\text{CF}_3$ ,  $\text{CHF}_2$ ,  $\text{CHCl}_2$ ,  $\text{CH}_2\text{F}$ ,  $\text{CH}_2\text{Cl}$ ,  $\text{CH}_2\text{OH}$ ,  $\text{CH}_2\text{OMe}$ ,  $\text{CH}_2\text{CN}$ , CN,  $\text{<CCH>}$ , 1-propynyl, 2-propynyl, vinyl, OH, OMe, OEt, cyclopropyl,  $\text{CF}_2\text{CF}_3$ ,  $\text{CH}=\text{CHCN}$ , allyl, azido, SCN, or  $\text{SCHF}_2$ ;  $\text{R}_2$  is H, Me, Et, n-Pr, i-Pr, formyl,  $\text{CF}_3$ ,  $\text{CHF}_2$ ,  $\text{CHCl}_2$ ,  $\text{CH}_2\text{F}$ ,  $\text{CH}_2\text{Cl}$ ,  $\text{CH}_2\text{OH}$ ,  $\text{CH}_2\text{OMe}$ ,  $\text{CH}_2\text{CN}$ , CN,  $\text{<CCH>}$ , 1-propynyl 2-propynyl, vinyl, Ac, F, Cl, OH, OMe, OEt, O-n-Pr, OAc,  $\text{NMe}_2$ ,  $\text{NEt}_2$ , SMe,  $\text{SEt}$ ,  $\text{SOCF}_3$ ,  $\text{OCF}_2\text{CF}_2\text{H}$ , COEt, cyclopropyl,  $\text{CF}_2\text{CF}_3$ ,  $\text{CH}=\text{CHCN}$ , allyl, azido,  $\text{OCF}_3$ ,  $\text{OCHF}_2$ , O-i-Pr, SCN,  $\text{SCHF}_2$ ,  $\text{SOMe}$ , NH-CN, or joined with  $\text{R}^3$  and the phenyl carbons to which  $\text{R}^2$  and  $\text{R}^3$  are attached to form an ethylenedioxy, a dihydrofuryl ring with the oxygen adjacent to a phenyl carbon, or a dihydropyrryl ring with the oxygen adjacent to a phenyl carbon;

25

$\text{R}^3$  is H, ET, or joined with  $\text{R}^2$  and the phenyl carbons to which  $\text{R}^2$  and  $\text{R}^3$  are attached to form an ethylenedioxy, a dihydrofuryl ring with the oxygen adjacent to a phenyl carbon, or a dihydropyrryl ring with the oxygen adjacent to a phenyl carbon;

30

$\text{R}^4$ ,  $\text{R}^5$ , and  $\text{R}^6$  are independently H, Me, Et, F, Cl, Br, formyl,  $\text{CF}_3$ ,  $\text{CHF}_2$ ,  $\text{CHCl}_2$ ,  $\text{CH}_2\text{F}$ ,  $\text{CH}_2\text{Cl}$ ,  $\text{CH}_2\text{OH}$ , CN,  $\text{C}=\text{CH}$ , 1-propynyl, 2-propynyl, vinyl, OMe, OEt, SMe, or SEt..

35

29. An isolated host cell into which the gene expression modulation system according to claim 1 or 7 has been introduced.

40

30. The isolated host cell according to claim 29, wherein the host cell is selected from the group consisting of a bacterial cell, a fungal cell, a yeast cell, a plant cell, an animal cell, and a mammalian cell.

45

31. The isolated host cell according to claim 30, wherein the mammalian cell is a murine cell or a human cell.

50

32. A non-human organism comprising a host cell into which the gene expression modulation system according to claim 1 or 7 has been introduced.

55

33. The non-human organism according to claim 32, wherein the non-human organism is selected from the group consisting of a bacterium, a fungus, a yeast, a plant, an animal, and a mammal.

60

34. The non-human organism according to claim 33, wherein the mammal is selected from the group consisting of a mouse, a rat, a rabbit, a cat, a dog, a bovine, a goat, a pig, a horse, a sheep, a monkey, and a chimpanzee.

#### Patentansprüche

1. Die Genexpression modulierendes System, welches folgendes umfaßt:

55

a) eine erste Genexpressionkassette, die in einer Wirtszelle exprimiert werden kann, umfassend eine Polynukleotidsequenz, die ein erstes Polypeptid codiert, welches folgendes umfaßt:

5 i) eine DNA-Bindungsdomäne, die ein Antwortelement, das mit einem Gen, dessen Expression moduliert werden soll, assoziiert ist, erkennt,  
 ii) eine Ligandenbindungsdomäne, die eine Ligandenbindungsdomäne eines Ecdysonrezeptorpolypeptids umfaßt;

b) eine zweite Genexpressionskassette, die in der Wirtszelle exprimiert werden kann, umfassend eine Polynukleotidsequenz, die ein zweites Polypeptid codiert, welches folgendes umfaßt:

10 i) eine Transaktivierungsdomäne und  
 ii) eine Ligandenbindungsdomäne, die eine Ligandenbindungsdomäne eines verkürzten Retinoid X-Rezeptors, der eine Deletion wenigstens der A- und der B-Domäne aufweist, umfaßt,

wobei die Transaktivierungsdomäne von einem Kernrezeptor stammt, der kein Ecdysonrezeptor, kein Retinoid X-Rezeptor oder kein Ultraspiracle-Rezeptor ist, und wobei die Ligandenbindungsdomänen des ersten Polypeptids und des zweiten Polypeptids verschieden sind und dimerisieren.

15 2. Die Genexpression modulierendes System nach Anspruch 1, welches weiterhin eine dritte Genexpressionskassette umfaßt, die folgendes umfaßt:

20 i) ein Antwortelement, an das die DNA-Bindungsdomäne des Polypeptids bindet,  
 ii) einen Promotor, der durch die Transaktivierungsdomäne des zweiten Polypeptids aktiviert wird, und  
 iii) das Gen, dessen Expression moduliert werden soll.

25 3. Die Genexpression modulierendes System nach Anspruch 1, wobei die Ligandenbindungsdomäne des ersten Polypeptids durch ein Polynukleotid codiert wird, welches eine Nukleinsäuresequenz umfaßt, die aus der Gruppe ausgewählt ist, bestehend aus SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9 und SEQ ID NO:10.

30 4. Die Genexpression modulierendes System nach Anspruch 1, wobei die Ligandenbindungsdomäne des ersten Polypeptids eine Aminosäuresequenz umfaßt, die aus der Gruppe ausgewählt ist, bestehend aus SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19 und SEQ ID NO:20.

35 5. Die Genexpression modulierendes System nach Anspruch 1, wobei die Ligandenbindungsdomäne des zweiten Polypeptids durch ein Polynukleotid codiert wird, welches eine Nukleinsäuresequenz umfaßt, die aus der Gruppe ausgewählt ist, bestehend aus SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29 und SEQ ID NO:30.

40 6. Die Genexpression modulierendes System nach Anspruch 1, wobei die Ligandenbindungsdomäne des zweiten Polypeptids eine Aminosäuresequenz umfaßt, die aus der Gruppe ausgewählt ist, bestehend aus SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39 und SEQ ID NO:40.

45 7. Die Genexpression modulierendes System, welches folgendes umfaßt:

a) eine erste Genexpressionskassette, die in einer Wirtszelle exprimiert werden kann, umfassend eine Polynukleotidsequenz, welche ein erstes Polypeptid codiert, das folgendes umfaßt:

50 i) eine DNA-Bindungsdomäne, die ein Antwortelement, das mit einem Gen, dessen Expression moduliert werden soll, assoziiert ist, erkennt und  
 ii) eine Ligandenbindungsdomäne, die eine Ligandenbindungsdomäne eines verkürzten Retinoid X-Rezeptors, der eine Deletion wenigstens der A- und der B-Domäne aufweist, umfaßt und

b) eine zweite Genexpressionskassette, die in einer Wirtszelle exprimiert werden kann, umfassend eine Polynukleotidsequenz, die ein zweites Polypeptid codiert, das folgendes umfaßt:

i) eine Transaktivierungsdomäne und  
 ii) eine Ligandenbindungsdomäne, die eine Ligandenbindungsdomäne eines Ecdysonrezeptors umfaßt,

wobei die Ligandenbindungsdomänen des ersten Polypeptids und des zweiten Polypeptids verschieden sind und dimerisieren.

5 8. Die Genexpression modulierendes System nach Anspruch 7, welches weiterhin eine dritte Genexpressionskassette umfaßt, die folgendes umfaßt:

- i) ein Antwortelement, an das die DNA-Bindungsdomäne des ersten Polypeptids bindet,
- ii) einen Promotor, der durch die Transaktivierungsdomäne des zweiten Polypeptids aktiviert wird, und
- iii) das Gen, dessen Expression moduliert werden soll.

10 9. Die Genexpression modulierendes System nach Anspruch 7, wobei die Ligandenbindungsdomäne des ersten Polypeptids durch ein Polynukleotid codiert wird, welches eine Nukleinsäuresequenz umfaßt, die aus der Gruppe ausgewählt ist, bestehend aus SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29 und SEQ ID NO:30.

15 10. Die Genexpression modulierendes System nach Anspruch 7, wobei die Ligandenbindungsdomäne des ersten Polypeptids eine Aminosäuresequenz umfaßt, die aus der Gruppe ausgewählt ist, bestehend aus SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39 und SEQ ID NO:40.

20 11. Die Genexpression modulierendes System nach Anspruch 7, wobei die Ligandenbindungsdomäne des zweiten Polypeptids durch ein Polynukleotid codiert wird, welches eine Nukleinsäuresequenz umfaßt, die aus der Gruppe ausgewählt ist, bestehend aus SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9 und SEQ ID NO:10.

25 12. Die Genexpression modulierendes System nach Anspruch 7, wobei die Ligandenbindungsdomäne des zweiten Polypeptids eine Aminosäuresequenz umfaßt, die aus der Gruppe ausgewählt ist, bestehend aus SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19 und SEQ ID NO:20.

30 13. Genexpressionskassette, welche ein Polynukleotid umfaßt, das ein hybrides Polypeptid codiert, ausgewählt aus der Gruppe, bestehend aus:

- a) einer DNA-Bindungsdomäne und einer Ecdysonrezeptor-Ligandenbindungsdomäne, wobei die DNA-Bindungsdomäne von einem Kernrezeptor stammt, der kein Ecdysonrezeptor ist,
- b) einer DNA-Bindungsdomäne und einer Ligandenbindungsdomäne eines verkürzten Retinoid X-Rezeptors, welcher eine Deletion wenigstens der A- und der B-Domäne aufweist, wobei die DNA-Bindungsdomäne von einem Kernrezeptor stammt, der kein Retinoid X-Rezeptor ist,
- c) einer Transaktivierungsdomäne und einer Ecdysonrezeptor-Ligandenbindungsdomäne, wobei die Transaktivierungsdomäne von einem Kernrezeptor stammt, der kein Ecdysonrezeptor ist, und
- d) einer Transaktivierungsdomäne und einer Ligandenbindungsdomäne eines verkürzten Retinoid X-Rezeptors, welcher eine Deletion wenigstens der A- und der B-Domäne aufweist, wobei die Transaktivierungsdomäne von einem Kernrezeptor stammt, der kein Retinoid X-Rezeptor ist.

45 14. Genexpressionskassette nach Anspruch 13, wobei die Expressionskassette ein Polynukleotid umfaßt, welches ein hybrides Polypeptid codiert, das folgendes aufweist:

- eine DNA-Bindungsdomäne und eine Ecdysonrezeptor-Ligandenbindungsdomäne, wobei die DNA-Bindungsdomäne von einem Kerrezeptor stammt, der kein Ecdysonrezeptor ist,
- eine DNA-Bindungsdomäne und eine Retinoid X-Rezeptor-Ligandenbindungsdomäne, wobei die DNA-Bindungsdomäne von einem Kernrezeptor stammt, der kein Retinoid X-Rezeptor ist, und

wobei die DNA-Bindungsdomäne eine GAL4-DNA-Bindungsdomäne oder eine LexA-DNA-Bindungsdomäne ist.

55 15. Genexpressionskassette nach Anspruch 13, wobei die Expressionskassette ein Polynukleotid umfaßt, welches ein hybrides Polypeptid codiert, das folgendes aufweist:

- eine Transaktivierungsdomäne und eine Ecdysonrezeptor-Ligandenbindungsdomäne, wobei die Transaktivie-

rungsdomäne von einem Kernrezeptor stammt, der kein Ecdysonrezeptor ist, oder eine Transaktivierungsdomäne und eine Retinoid X-Rezeptor-Ligandenbindungsdomäne, wobei die Transaktivierungsdomäne von einem Kernrezeptor stammt, der kein Retinoid X-Rezeptor ist,

5 und wobei die Transaktivierungsdomäne eine VP16-Transaktivierungsdomäne ist.

16. Genexpressionskassette, welche ein Polynukleotid umfaßt, das ein hybrides Polypeptid codiert, welches eine DNA-Bindungsdomäne umfaßt, die durch ein Polynukleotid codiert ist, welches eine Nukleinsäuresequenz umfaßt, die aus der Gruppe ausgewählt ist, bestehend aus GAL4 DBD (SEQ ID NO:41) oder LexA DBD (SEQ ID NO:43), und eine Ecdysonrezeptor-Ligandenbindungsdomäne umfaßt, die durch ein Polynukleotid codiert ist, welches eine Nukleinsäuresequenz umfaßt, die aus der Gruppe ausgewählt ist, bestehend aus SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9 und SEQ ID NO:10.
17. Genexpressionskassette, welche ein Polynukleotid umfaßt, das ein hybrides Polypeptid codiert, welches eine DNA-Bindungsdomäne umfaßt, die eine Aminosäuresequenz umfaßt, ausgewählt aus der Gruppe, bestehend aus GAL4 DBD (SEQ ID NO:42) oder LexA DBD (SEQ ID NO:44), umfaßt, und eine Ecdysonrezeptor-Ligandenbindungsdomäne, die eine Aminosäuresequenz umfaßt, ausgewählt aus der Gruppe, bestehend aus SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19 und SEQ ID NO:20.
18. Genexpressionskassette, welche ein Polynukleotid umfaßt, das ein hybrides Polypeptid codiert, welches eine DNA-Bindungsdomäne umfaßt, die durch ein Polynukleotid codiert wird, welches eine Nukleinsäuresequenz umfaßt, ausgewählt aus der Gruppe, bestehend aus GAL4 DBD (SEQ ID NO:41) oder LexA DBD (SEQ ID NO:43), und eine Retinoid X-Rezeptor-Ligandenbindungsdomäne umfaßt, die durch ein Polynukleotid codiert wird, welches eine Nukleinsäuresequenz umfaßt, ausgewählt aus der Gruppe, bestehend aus SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29 und SEQ ID NO:30.
19. Genexpressionskassette, welche ein Polynukleotid umfaßt, das ein hybrides Polypeptid codiert, welches eine DNA-Bindungsdomäne umfaßt, die eine Aminosäuresequenz umfaßt, ausgewählt aus der Gruppe, bestehend aus GAL4 DBD (SEQ ID NO:42) oder LexA DBD (SEQ ID NO:44), und eine Retinoid X-Rezeptor-Ligandenbindungsdomäne umfaßt, die eine Aminosäuresequenz umfaßt, ausgewählt aus der Gruppe, bestehend aus SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39 und SEQ ID NO:40.
20. Genexpressionskassette, welche ein Polynukleotid umfaßt, das ein hybrides Polypeptid codiert, welches eine Transaktivierungsdomäne umfaßt, die durch ein Polynukleotid codiert wird, welches eine Nukleinsäuresequenz der SEQ ID NO:45 umfaßt, und eine Ecdysonrezeptor-Ligandenbindungsdomäne umfaßt, die durch ein Polynukleotid codiert wird, welches eine Nukleinsäuresequenz umfaßt, ausgewählt aus der Gruppe, bestehend aus SEQ ID NO:1, SEQ ID NO:2, SEQ ID NO:3, SEQ ID NO:4, SEQ ID NO:5, SEQ ID NO:6, SEQ ID NO:7, SEQ ID NO:8, SEQ ID NO:9 und SEQ ID NO:10.
21. Genexpressionskassette, welche ein Polynukleotid umfaßt, das ein hybrides Polypeptid codiert, welches eine Transaktivierungsdomäne umfaßt, die eine Aminosäuresequenz der SEQ ID NO:46 umfaßt, und eine Ecdysonrezeptor-Ligandenbindungsdomäne umfaßt, die eine Aminosäuresequenz umfaßt, ausgewählt aus der Gruppe, bestehend aus SEQ ID NO:11, SEQ ID NO:12, SEQ ID NO:13, SEQ ID NO:14, SEQ ID NO:15, SEQ ID NO:16, SEQ ID NO:17, SEQ ID NO:18, SEQ ID NO:19 und SEQ ID NO:20.
22. Genexpressionskassette, welche ein Polynukleotid umfaßt, das ein hybrides Polypeptid codiert, welches eine Transaktivierungsdomäne umfaßt, die durch ein Polynukleotid codiert wird, welches eine Nukleinsäuresequenz der SEQ ID NO:45 umfaßt, und eine Retinoid X-Rezeptor-Ligandenbindungsdomäne umfaßt, die durch ein Polynukleotid codiert wird, welches eine Nukleinsäuresequenz umfaßt, ausgewählt aus der Gruppe, bestehend aus SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29 und SEQ ID NO:30.
23. Genexpressionskassette, welche ein Polynukleotid umfaßt, das ein hybrides Polypeptid codiert, welches eine Transaktivierungsdomäne umfaßt, die eine Aminosäuresequenz der SEQ ID NO:46 umfaßt, und eine Retinoid X-Rezeptor-Ligandenbindungsdomäne umfaßt, die eine Aminosäuresequenz umfaßt, ausgewählt aus der Gruppe, bestehend

aus SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39 und SEQ ID NO:40.

5 24. Isoliertes Polypeptid, welches ein verkürztes Retinoid X-Rezeptorpolypeptid codiert, wobei das Polynukleotid eine Nukleinsäuresequenz umfaßt, ausgewählt aus der Gruppe, bestehend aus SEQ ID NO:21, SEQ ID NO:22, SEQ ID NO:23, SEQ ID NO:24, SEQ ID NO:25, SEQ ID NO:26, SEQ ID NO:27, SEQ ID NO:28, SEQ ID NO:29 und SEQ ID NO:30.

10 25. Isoliertes verkürztes Retinoid X-Rezeptorpolypeptid, welches eine Aminosäuresequenz umfaßt, ausgewählt aus der Gruppe, bestehend aus SEQ ID NO:31, SEQ ID NO:32, SEQ ID NO:33, SEQ ID NO:34, SEQ ID NO:35, SEQ ID NO:36, SEQ ID NO:37, SEQ ID NO:38, SEQ ID NO:39 und SEQ ID NO:40.

15 26. Isoliertes Polynukleotid, welches eine Polynukleotidsequenz umfaßt, die das isolierte Polypeptid nach Anspruch 25 codiert.

27. Verfahren zum Modulieren der Expression eines Gens in einer Wirtszelle, die das zu modulierende Gen enthält, welches die folgenden Stufen umfaßt:

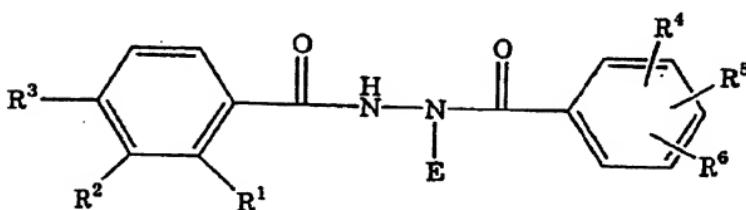
20 a) Einbringen des die Genexpression modulierenden Systems nach einem der Ansprüche 1 oder 2 in die Wirtszelle und  
b) Einbringen eines Liganden, der sich unabhängig mit den Ligandenbindungsdomänen des ersten Polypeptids und des zweiten Polypeptids vereinigt, in die Wirtszelle,

wobei das zu exprimierende Gen eine Komponente eines chimären Gens ist, welches folgendes umfaßt:

25 i) ein Antwortelement, an das die DNA-Bindungsdomäne des ersten Polypeptids bindet,  
ii) einen Promotor, der durch die Transaktivierungsdomäne des zweiten Polypeptids aktiviert wird, und  
iii) das Gen, dessen Expression moduliert werden soll,

30 wobei ein Komplex gebildet wird, der den Liganden, das erste Polypeptid und das zweite Polypeptid umfaßt, und wobei der Komplex die Expression des Gens in der Wirtszelle moduliert.

28. Verfahren nach Anspruch 27, wobei der Ligand eine Verbindung der Formel



ist, wobei:

50 E ein (C<sub>4</sub>-C<sub>8</sub>)-Alkyl, das ein tertiäres Kohlenstoffatom enthält, oder ein Cyano-(C<sub>3</sub>-C<sub>5</sub>)-alkyl, das ein tertiäres Kohlenstoffatom enthält, ist,  
R<sup>1</sup> H, Me, Et, i-Pr, F, Formyl, CF<sub>3</sub>, CHF<sub>2</sub>, CHCl<sub>2</sub>, CH<sub>2</sub>F, CH<sub>2</sub>Cl, CH<sub>2</sub>OH, CH<sub>2</sub>OMe, CH<sub>2</sub>CN, CN, CCH, 1-Propinyl, 2-Propinyl, Vinyl, OH, OMe, OEt, Cyclopropyl, CF<sub>2</sub>CF<sub>3</sub>, CH=CHCN, Allyl, Azido, SCN oder SCHF<sub>2</sub> ist,  
R<sup>2</sup> H, Me, Et, n-Pr, i-Pr, Formyl, CF<sub>3</sub>, CHF<sub>2</sub>, CHCl<sub>2</sub>, CH<sub>2</sub>F, CH<sub>2</sub>Cl, CH<sub>2</sub>OH, CH<sub>2</sub>OMe, CH<sub>2</sub>CN, CN, CCH, 1-Propinyl, 2-Propinyl, Vinyl, Ac, F, Cl, OH, OMe, OEt, O-n-Pr, OAC, NMe<sub>2</sub>, NEt<sub>2</sub>, SME, SET, SOCF<sub>3</sub>, OCFC<sub>2</sub>H, COEt, Cyclopropyl, CF<sub>2</sub>CF<sub>3</sub>, CH=CHCN, Allyl, Azido, OCF<sub>3</sub>, OCHF<sub>2</sub>, O-i-Pr, SCN, SCHF<sub>2</sub>, SOMe, NH-CN ist oder mit R<sup>3</sup> und den Phenylkohlenstoffen, an die R<sup>2</sup> und R<sup>3</sup> angebunden sind, unter Bildung von Ethylenoxyd, eines Dihydrofuryrlings, bei dem der Sauerstoff zu einem Phenylkohlenstoff benachbart ist, oder eines Dihy-

dropyrlings, bei dem der Sauerstoff zu einem Phenylkohlenstoff benachbart ist, verbunden ist, R<sup>3</sup> H, Et oder mit R<sup>2</sup> und den Phenylkohlenstoffen, an die R<sup>2</sup> und R<sup>3</sup> angebunden sind, unter Bildung von Ethyldioxy, eines Dihydrofurylring, bei dem der Sauerstoff zu einem Phenylkohlenstoff benachbart ist, oder eines Dihydropyrrylring, bei dem der Sauerstoff zu einem Phenylkohlenstoff benachbart ist, verbunden ist, R<sup>4</sup>, R<sup>5</sup> und R<sup>6</sup> unabhängig voneinander H, Me, Et, F, Cl, Br, Formyl, CF<sub>3</sub>, CHF<sub>2</sub>, CHCl<sub>2</sub>, CH<sub>2</sub>F, CH<sub>2</sub>Cl, CH<sub>2</sub>OH, CN, C°CH, 1-Propinyl, 2-Propinyl, Vinyl, OMe, OEt, SMe oder SEt sind.

5 29. Isolierte Wirtszelle, in die das die Genexpression modulierende System nach Anspruch 1 oder 7 eingebracht wurde.

10 30. Isolierte Wirtszelle nach Anspruch 29, wobei die Wirtszelle aus der Gruppe ausgewählt ist, bestehend aus einer Bakterienzelle, einer Pilzelle, einer Hefezelle, einer Pflanzenzelle, einer Tierzelle und einer Säugerzelle.

15 31. Isolierte Wirtszelle nach Anspruch 30, wobei die Säugerzelle eine Mauszelle oder eine menschliche Zelle ist.

20 32. Nicht-menschlicher Organismus, welcher eine Wirtszelle, in die das die Genexpression modulierende System nach Anspruch 1 oder 7 eingebracht wurde, umfaßt.

33. Nicht-menschlicher Organismus nach Anspruch 32, wobei der nicht-menschliche Organismus aus der Gruppe ausgewählt ist, bestehend aus einem Bakterium, einem Pilz, einer Hefe, einer Pflanze, einem Tier und einem Säuger.

25 34. Nicht-menschlicher Organismus nach Anspruch 33, wobei der Säuger aus der Gruppe ausgewählt ist, bestehend aus einer Maus, einer Ratte, einem Kaninchen, einer Katze, einem Hund, einem Rind, einer Ziege, einem Schwein, einem Pferd, einem Schaf, einem Affen und einem Schimpanse.

#### Revendications

1. Système de modulation d'expression génique comprenant :

30 a) une première cassette d'expression génique qui est capable d'être exprimée dans une cellule hôte comprenant une séquence polynucléotidique qui code pour un premier polypeptide comprenant :  
 i) un domaine de liaison à l'ADN qui reconnaît un élément de réponse associé à un gène dont l'expression doit être modulée ;  
 ii) un domaine de liaison au ligand comprenant un domaine de liaison au ligand issu d'un polypeptide du récepteur de l'ecdysone ;

35 b) une deuxième cassette d'expression génique qui est capable d'être exprimée dans la cellule hôte comprenant une séquence polynucléotidique qui code pour un second polypeptide comprenant :

40 i) un domaine de transactivation ; et  
 ii) un domaine de liaison au ligand comprenant un domaine de liaison au ligand issu d'un récepteur X des rétinoides tronqué comprenant une délétion d'au moins les domaines A et B ;

45 dans lequel le domaine de transactivation provient d'un récepteur nucléaire autre qu'un récepteur de l'ecdysone, d'un récepteur X des rétinoides ou d'un récepteur ultraspiracle ; et dans lequel les domaines de liaison au ligand issus du premier polypeptide et du second polypeptide sont différents et se dimérisent.

50 2. Système de modulation d'expression génique selon la revendication 1, comprenant de plus une troisième cassette d'expression génique comprenant :

i) un élément de réponse auquel se lie le domaine de liaison à l'ADN du polypeptide ;  
 ii) un promoteur qui est activé par le domaine de transactivation du second polypeptide ; et  
 iii) le gène dont l'expression doit être modulée.

55 3. Système d'expression génique selon la revendication 1, dans lequel le domaine de liaison au ligand du premier polypeptide est codé par un polynucléotide comprenant une séquence d'acide nucléique choisie dans le groupe formé par SEQ ID N° 1, SEQ ID N° 2, SEQ ID N° 3, SEQ ID N° 4, SEQ ID N° 5, SEQ ID N° 6, SEQ ID N° 7, SEQ

ID N° 8, SEQ ID N° 9 et SEQ ID N° 10.

4. Système de modulation d'expression génique selon la revendication 1, dans lequel le domaine de liaison au ligand du premier polypeptide comprend une séquence d'acides aminés choisie dans le groupe formé par SEQ ID N° 11, SEQ ID N° 12, SEQ ID N° 13, SEQ ID N° 14, SEQ ID N° 15, SEQ ID N° 16, SEQ ID N° 17, SEQ ID N° 18, SEQ ID N° 19 et SEQ ID N° 20.
5. Système de modulation d'expression génique selon la revendication 1, dans lequel le domaine de liaison au ligand du second polypeptide est codé par un polynucléotide comprenant une séquence d'acide nucléique choisie dans le groupe formé par SEQ ID N° 21, SEQ ID N° 22, SEQ ID N° 23, SEQ ID N° 24, SEQ ID N° 25, SEQ ID N° 26, SEQ ID N° 27, SEQ ID N° 28, SEQ ID N° 29 et SEQ ID N° 30.
10. Système de modulation d'expression génique selon la revendication 1, dans lequel le domaine de liaison au ligand du second polypeptide comprend une séquence d'acides aminés choisie dans le groupe formé par SEQ ID N° 31, SEQ ID N° 32, SEQ ID N° 33, SEQ ID N° 34, SEQ ID N° 35, SEQ ID N° 36, SEQ ID N° 37, SEQ ID N° 38, SEQ ID N° 39 et SEQ ID N° 40.
15. Système de modulation d'expression génique selon la revendication 1, dans lequel le domaine de liaison au ligand du second polypeptide comprend une séquence d'acides aminés choisie dans le groupe formé par SEQ ID N° 31, SEQ ID N° 32, SEQ ID N° 33, SEQ ID N° 34, SEQ ID N° 35, SEQ ID N° 36, SEQ ID N° 37, SEQ ID N° 38, SEQ ID N° 39 et SEQ ID N° 40.
20. Système de modulation d'expression génique comprenant :
- a) une première cassette d'expression génique qui est capable d'être exprimée dans une cellule hôte comprenant une séquence polynucléotidique qui code pour un premier polypeptide comprenant :
  - i) un domaine de liaison à l'ADN qui reconnaît un élément de réponse associé à un gène dont l'expression doit être modulée ; et
  25. ii) un domaine de liaison au ligand comprenant un domaine de liaison au ligand issu d'un récepteur X des rétinoïdes tronqué comprenant une délétion d'au moins les domaines A et B ; et
- b) une deuxième cassette d'expression génique qui est capable d'être exprimée dans une cellule hôte comprenant une séquence polynucléotidique qui code pour un second polypeptide comprenant :
  - i) un domaine de transactivation ; et
  30. ii) un domaine de liaison au ligand comprenant un domaine de liaison au ligand issu d'un récepteur de l'ecdysone ;
35. dans lequel les domaines de liaison au ligand issus du premier polypeptide et du second polypeptide sont différents et se dimérisent.
8. Système de modulation d'expression génique selon la revendication 7, comprenant de plus une troisième cassette d'expression génique comprenant :
40. i) un élément de réponse auquel se lie le domaine de liaison à l'ADN du premier polypeptide ;
- ii) un promoteur qui est activé par le domaine de transactivation du second polypeptide ; et
- iii) le gène dont l'expression est modulée.
45. 9. Système de modulation d'expression génique selon la revendication 7, dans lequel le domaine de liaison au ligand du premier polypeptide est codé par un polynucléotide comprenant une séquence d'acide nucléique choisie dans le groupe formé par SEQ ID N° 21, SEQ ID N° 22, SEQ ID N° 23, SEQ ID N° 24, SEQ ID N° 25, SEQ ID N° 26, SEQ ID N° 27, SEQ ID N° 28, SEQ ID N° 29 et SEQ ID N° 30.
50. 10. Système de modulation d'expression génique selon la revendication 7, dans lequel le domaine de liaison au ligand du premier polypeptide comprend une séquence d'acides aminés choisie dans le groupe formé par SEQ ID N° 31, SEQ ID N° 32, SEQ ID N° 33, SEQ ID N° 34, SEQ ID N° 35, SEQ ID N° 36, SEQ ID N° 37, SEQ ID N° 38, SEQ ID N° 39 et SEQ ID N° 40.
55. 11. Système de modulation d'expression génique selon la revendication 7, dans lequel le domaine de liaison au ligand du second polypeptide est codé par un polynucléotide comprenant une séquence d'acide nucléique choisie dans le groupe formé par SEQ ID N° 1, SEQ ID N° 2, SEQ ID N° 3, SEQ ID N° 4, SEQ ID N° 5, SEQ ID N° 6, SEQ ID N° 7, SEQ ID N° 8, SEQ ID N° 9 et SEQ ID N° 10.

12. Système de modulation d'expression génique selon la revendication 7, dans lequel le domaine de liaison au ligand du second polypeptide comprend une séquence d'acides aminés choisie dans le groupe formé par SEQ ID N° 11, SEQ ID N° 12, SEQ ID N° 13, SEQ ID N° 14, SEQ ID N° 15, SEQ ID N° 16, SEQ ID N° 17, SEQ ID N° 18, SEQ ID N° 19 et SEQ ID N° 20.

13. Cassette d'expression génique comprenant un polynucléotide codant pour un polypeptide hybride choisi dans le groupe formé par :

- a) un domaine de liaison à l'ADN et un domaine de liaison au ligand de récepteur de l'ecdysone, le domaine de liaison à l'ADN étant issu d'un récepteur nucléaire autre qu'un récepteur de l'ecdysone ;
- b) un domaine de liaison à l'ADN et un domaine de liaison au ligand issu d'un récepteur X des rétinoïdes tronqué comprenant une déletion d'au moins les domaines A et B, le domaine de liaison à l'ADN étant issu d'un récepteur nucléaire autre qu'un récepteur X des rétinoïdes ;
- c) un domaine de transactivation et un domaine de liaison au ligand de récepteur de l'ecdysone, le domaine de transactivation étant issu d'un récepteur nucléaire autre qu'un récepteur de l'ecdysone ; et
- d) un domaine de transactivation et un domaine de liaison au ligand issu d'un récepteur X des rétinoïdes tronqué comprenant une déletion d'au moins les domaines A et B, le domaine de transactivation étant issu d'un récepteur nucléaire autre qu'un récepteur X des rétinoïdes.

14. Cassette d'expression génique selon la revendication 13, dans laquelle ladite cassette d'expression génique comprend un polynucléotide codant pour un polypeptide hybride comprenant :

- un domaine de liaison à l'ADN et un domaine de liaison au ligand de récepteur de l'ecdysone, le domaine de liaison à l'ADN étant issu d'un récepteur nucléaire autre qu'un récepteur de l'ecdysone ; ou
- un domaine de liaison à l'ADN et un domaine de liaison au ligand de récepteur X des rétinoïdes, le domaine de liaison à l'ADN étant issu d'un récepteur nucléaire autre qu'un récepteur X des rétinoïdes ; et

dans laquelle le domaine de liaison à l'ADN est un domaine de liaison à l'ADN de GAL4 ou un domaine de liaison à l'ADN de LexA.

15. Cassette d'expression génique selon la revendication 13, dans laquelle ladite cassette d'expression génique comprend un polynucléotide codant pour un polypeptide hybride comprenant :

- un domaine de transactivation et un domaine de liaison au ligand de récepteur de l'ecdysone, le domaine de transactivation étant issu d'un récepteur nucléaire autre qu'un récepteur de l'ecdysone ;
- ou un domaine de transactivation et un domaine de liaison au ligand de récepteur X des rétinoïdes, le domaine de transactivation étant issu d'un récepteur nucléaire autre qu'un récepteur X des rétinoïdes ;

et dans laquelle le domaine de transactivation est un domaine de transactivation de VP16.

16. Cassette d'expression génique comprenant un polynucléotide codant pour un polypeptide hybride comprenant un domaine de liaison à l'ADN codé par un polynucléotide comprenant une séquence d'acide nucléique choisie dans le groupe formé par un DBD de GAL4 (SEQ ID N° 41) ou un DBD de LexA (SEQ ID N° 43) et un domaine de liaison au ligand de récepteur de l'ecdysone codé par un polynucléotide comprenant une séquence d'acide nucléique choisie dans le groupe formé par SEQ ID N° 1, SEQ ID N° 2, SEQ ID N° 3, SEQ ID N° 4, SEQ ID N° 5, SEQ ID N° 6, SEQ ID N° 7, SEQ ID N° 8, SEQ ID N° 9 et SEQ ID N° 10.

17. Cassette d'expression génique comprenant un polynucléotide codant pour un polypeptide hybride comprenant un domaine de liaison à l'ADN comprenant une séquence d'acides aminés choisie dans le groupe formé par un DBD de GAL4 (SEQ ID N° 42) ou un DBD de LexA (SEQ ID N° 44) et un domaine de liaison au ligand de récepteur de l'ecdysone comprenant une séquence d'acides aminés choisie dans le groupe formé par SEQ ID N° 11, SEQ ID N° 12, SEQ ID N° 13, SEQ ID N° 14, SEQ ID N° 15, SEQ ID N° 16, SEQ ID N° 17, SEQ ID N° 18, SEQ ID N° 19 et SEQ ID N° 20.

18. Cassette d'expression génique comprenant un polynucléotide codant pour un polypeptide hybride comprenant un domaine de liaison à l'ADN codé par un polynucléotide comprenant une séquence d'acide nucléique choisie dans le groupe formé par un DBD de GAL4 (SEQ ID N° 41) ou un DBD de LexA (SEQ ID N° 43) et un domaine de liaison au ligand de récepteur X des rétinoïdes codé par un polynucléotide comprenant une séquence d'acide nucléique

choisie dans le groupe formé par SEQ ID N° 21, SEQ ID N° 22, SEQ ID N° 23, SEQ ID N° 24, SEQ ID N° 25, SEQ ID N° 26, SEQ ID N° 27, SEQ ID N° 28, SEQ ID N° 29 et SEQ ID N° 30.

5        19. Cassette d'expression génique comprenant un polynucléotide codant pour un polypeptide hybride comprenant un domaine de liaison à l'ADN comprenant une séquence d'acides aminés choisie dans le groupe formé par un DBD de GAL4 (SEQ ID N° 42) ou un DBD de LexA (SEQ ID N° 44) et un domaine de liaison au ligand de récepteur X des rétinoïdes comprenant une séquence d'acides aminés choisie dans le groupe formé par SEQ ID N° 31, SEQ ID N° 32, SEQ ID N° 33, SEQ ID N° 34, SEQ ID N° 35, SEQ ID N° 36, SEQ ID N° 37, SEQ ID N° 38, SEQ ID N° 39 et SEQ ID N° 40.

10      20. Cassette d'expression génique comprenant un polynucléotide codant pour un polypeptide hybride comprenant un domaine de transactivation codé par un polynucléotide comprenant une séquence d'acide nucléique de SEQ ID N° 45 et un domaine de liaison au ligand de récepteur de l'ecdysone codé par un polynucléotide comprenant une séquence d'acide nucléique choisie dans le groupe formé par SEQ ID N° 1, SEQ ID N° 2, SEQ ID N° 3, SEQ ID N° 4, SEQ ID N° 5, SEQ ID N° 6, SEQ ID N° 7, SEQ ID N° 8, SEQ ID N° 9 et SEQ ID N° 10.

15      21. Cassette d'expression génique comprenant un polynucléotide codant pour un polypeptide hybride comprenant un domaine de transactivation comprenant une séquence d'acides aminés de SEQ ID N° 46 et un domaine de liaison au ligand de récepteur de l'ecdysone comprenant une séquence d'acides aminés choisie dans le groupe formé par SEQ ID N° 11, SEQ ID N° 12, SEQ ID N° 13, SEQ ID N° 14, SEQ ID N° 15, SEQ ID N° 16, SEQ ID N° 17, SEQ ID N° 18, SEQ ID N° 19 et SEQ ID N° 20.

20      22. Cassette d'expression génique comprenant un polynucléotide codant pour un polypeptide hybride comprenant un domaine de transactivation codé par un polynucléotide comprenant une séquence d'acide nucléique de SEQ ID N° 45 et un domaine de liaison au ligand de récepteur X des rétinoïdes codé par un polynucléotide comprenant une séquence d'acide nucléique choisie dans le groupe formé par SEQ ID N° 21, SEQ ID N° 22, SEQ ID N° 23, SEQ ID N° 24, SEQ ID N° 25, SEQ ID N° 26, SEQ ID N° 27, SEQ ID N° 28, SEQ ID N° 29 et SEQ ID N° 30.

25      23. Cassette d'expression génique comprenant un polynucléotide codant pour un polypeptide hybride comprenant un domaine de transactivation comprenant une séquence d'acides aminés de SEQ ID N° 46 et un domaine de liaison au ligand de récepteur X des rétinoïdes comprenant une séquence d'acides aminés choisie dans le groupe formé par SEQ ID N° 31, SEQ ID N° 32, SEQ ID N° 33, SEQ ID N° 34, SEQ ID N° 35, SEQ ID N° 36, SEQ ID N° 37, SEQ ID N° 38, SEQ ID N° 39 et SEQ ID N° 40.

30      24. Polynucléotide isolé codant pour un polypeptide de récepteur X des rétinoïdes tronqué, dans lequel le polynucléotide comprend une séquence d'acide nucléique choisie dans le groupe formé par SEQ ID N° 21, SEQ ID N° 22, SEQ ID N° 23, SEQ ID N° 24, SEQ ID N° 25, SEQ ID N° 26, SEQ ID N° 27, SEQ ID N° 28, SEQ ID N° 29 et SEQ ID N° 30.

35      25. Polypeptide de récepteur X des rétinoïdes tronqué isolé comprenant une séquence d'acides aminés choisie dans le groupe formé par SEQ ID N° 31, SEQ ID N° 32, SEQ ID N° 33, SEQ ID N° 34, SEQ ID N° 35, SEQ ID N° 36, SEQ ID N° 37, SEQ ID N° 38, SEQ ID N° 39 et SEQ ID N° 40.

40      26. Polynucléotide isolé comprenant une séquence polynucléotidique qui code pour le polypeptide isolé selon la revendication 25.

45      27. Procédé de modulation de l'expression d'un gène dans une cellule hôte comprenant le gène à moduler, comprenant les étapes de :

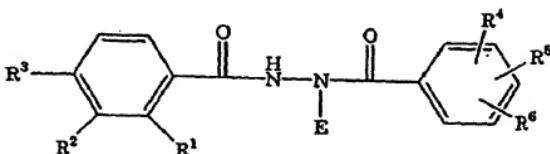
50            a) introduction dans la cellule hôte du système de modulation d'expression génique selon la revendication 1 ou 2 ; et  
                b) introduction dans la cellule hôte d'un ligand qui se combine indépendamment avec les domaines de liaison au ligand du premier polypeptide et du second polypeptide ;

55      dans lequel le gène à exprimer est un composant d'un gène chimère comprenant :  
                i) un élément de réponse auquel se lie le domaine de liaison à l'ADN issu du premier polypeptide ;  
                ii) un promoteur qui est activé par le domaine de transactivation du second polypeptide ; et  
                iii) un gène dont l'expression doit être modulée,

en sorte que soit formé un complexe comprenant le ligand, le premier polypeptide et le second polypeptide, et en sorte que le complexe module l'expression du gène dans la cellule hôte.

28. Procédé selon la revendication 27, dans lequel le ligand est un composé de formule :

5



10

15

dans laquelle :

$\text{E}$  est un groupe alkyle en  $\text{C}_4$  à  $\text{C}_6$  contenant un atome de carbone tertiaire ou un groupe cyano(alkyle en  $\text{C}_3$  à  $\text{C}_5$ ) contenant un atome de carbone tertiaire ;

$\text{R}^1$  est H, Me, Et, i-Pr, F, un groupe formyle,  $\text{CF}_3$ ,  $\text{CHF}_2$ ,  $\text{CHCl}_2$ ,  $\text{CH}_2\text{F}$ ,  $\text{CH}_2\text{Cl}$ ,  $\text{CH}_2\text{OH}$ ,  $\text{CH}_2\text{OMe}$ ,  $\text{CH}_2\text{CN}$ , CN, CCH, 1-propynyle, 2-propynyle, vinyle, OH, OMe, OEt, cyclopropyle,  $\text{CF}_2\text{CF}_3$ ,  $\text{CH}=\text{CHCN}$ , allyle, azido, SCN ou  $\text{SCHF}_2$  ;

$\text{R}^2$  est H, Me, Et, n-Pr, i-Pr, un groupe formyle,  $\text{CF}_3$ ,  $\text{CHF}_2$ ,  $\text{CHCl}_2$ ,  $\text{CH}_2\text{F}$ ,  $\text{CH}_2\text{Cl}$ ,  $\text{CH}_2\text{OH}$ ,  $\text{CH}_2\text{OMe}$ ,  $\text{CH}_2\text{CN}$ , CN, CCH, 1-propynyle, 2-propynyle, vinyle, Ac, F, Cl, OH, OMe, OEt, O-n-Pr, OAc,  $\text{NM}_2$ ,  $\text{NEt}_2$ , SMe, SET,  $\text{SOCF}_3$ ,  $\text{OCF}_2\text{CF}_3\text{H}$ , COEt, cyclopropyle,  $\text{CF}_2\text{CF}_3$ ,  $\text{CH}=\text{CHCN}$ , allyle, azido,  $\text{OCF}_3$ ,  $\text{OCHF}_2$ , O-i-Pr, SCN,  $\text{SCHF}_2$ , SOMe, NH-CN, ou est joint à  $\text{R}^3$  et aux atomes de carbone phényles auxquels  $\text{R}^2$  et  $\text{R}^3$  sont attachés pour former un groupe éthylénedioxy, un cycle dihydrofuryle dont l'atome d'oxygène est adjacent à un atome de carbone phényle, ou un cycle dihydropyryle dont l'atome d'oxygène est adjacent à un atome de carbone phényle ;

$\text{R}^3$  est H, Et, ou est joint à  $\text{R}^2$  et aux atomes de carbone phényles auxquels  $\text{R}^2$  et  $\text{R}^3$  sont attachés pour former un groupe éthylénedioxy, un cycle dihydrofuryle dont l'oxygène est adjacent à un atome de carbone phényle, ou un cycle dihydropyryle dont l'oxygène est adjacent à un atome de carbone phényle ;

$\text{R}^4$ ,  $\text{R}^5$  et  $\text{R}^6$  sont indépendamment H, Me, Et, F, Cl, Br, un groupe formyle,  $\text{CF}_3$ ,  $\text{CHF}_2$ ,  $\text{CHCl}_2$ ,  $\text{CH}_2\text{F}$ ,  $\text{CH}_2\text{Cl}$ ,  $\text{CH}_2\text{OH}$ , CN,  $\text{C}^=\text{CH}$ , 1-propynyle, 2-propynyle, vinyle, OMe, OEt, SMe ou SET.

29. Cellule hôte isolée dans laquelle a été introduit le système de modulation d'expression génique selon la revendication 1 ou 7.

40 30. Cellule hôte isolée selon la revendication 29, dans laquelle la cellule hôte est choisie dans le groupe formé par une cellule bactérienne, une cellule fongique, une cellule de levure, une cellule végétale, une cellule animale et une cellule de mammifère.

45 31. Cellule hôte isolée selon la revendication 30, dans laquelle la cellule de mammifère est une cellule murine ou une cellule humaine.

32. Organisme non humain comprenant une cellule hôte dans laquelle a été introduit le système de modulation d'expression génique selon la revendication 1 ou 7.

50 33. Organisme non humain selon la revendication 32, dans lequel l'organisme non humain est choisi dans le groupe formé par une bactérie, un champignon, une levure, une plante, un animal et un mammifère.

55 34. Organisme non humain selon la revendication 33, dans lequel le mammifère est choisi dans le groupe formé par une souris, un rat, un lapin, un chat, un chien, un bovin, une chèvre, un porc, un cheval, un mouton, un singe et un chimpanzé.

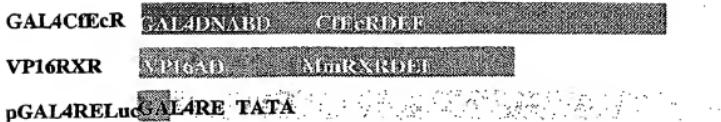


Figure 1

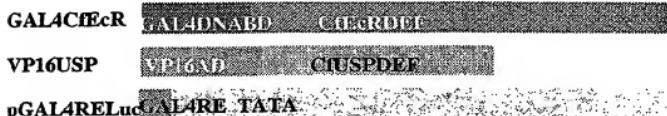


Figure 2



Figure 3

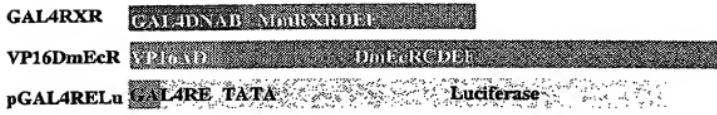


Figure 4

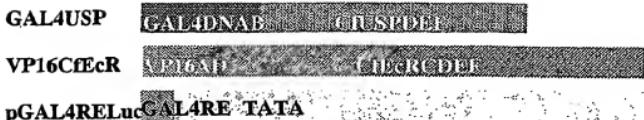


Figure 5

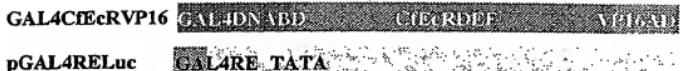


Figure 6

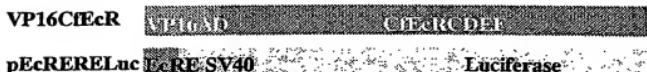


Figure 7

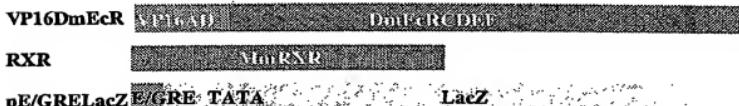


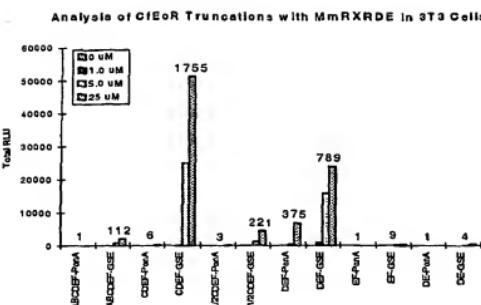
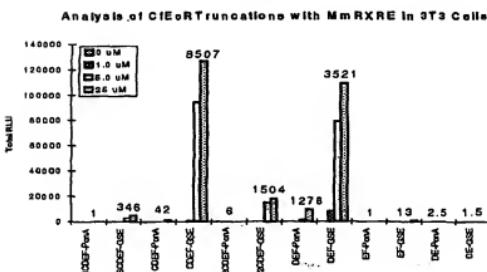
Figure 8



Figure 9



Figure 10

**Figure 11****Figure 12**

## Analysis of MmRXR Truncations with CfEcRCDEF in 3T3 Cells

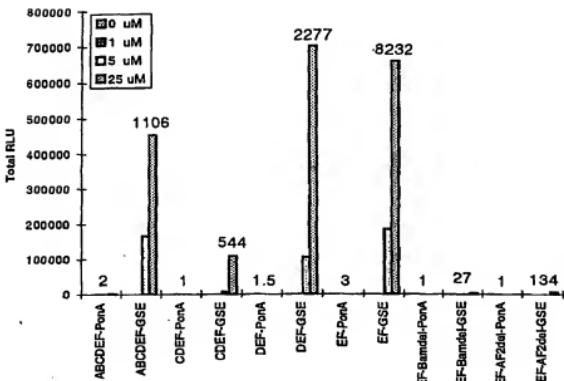


Figure 13

## Analysis of MmRXR Truncations with CfEcRCDEF in 3T3 Cells

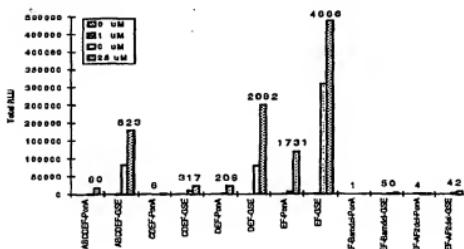


Figure 14

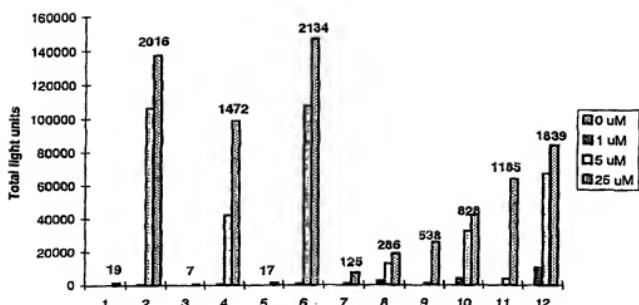


Figure 15